PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/32, C12N 15/62, 15/75 // 1/21, C12R 1/07

A1

(11) International Publication Number: **WO 99/04007**

(43) International Publication Date: 28 January 1999 (28.01.99)

(21) International Application Number: PCT/US98/14786

(22) International Filing Date: 16 July 1998 (16.07.98)

(30) Priority Data:

97305286.3	16 July 1997 (16.07.97)	EP
97305344.0	17 July 1997 (17.07.97)	EP

(71) Applicants *(for all designated States except US)*: GENENCOR INTERNATIONAL, INC. [US/US]; 4 Cambridge Place, 1870 South Winton Road, Rochester, NY 14618 (US). GENENCOR INTERNATIONAL B.V. [NL/NL]; Verrijn Stuartlaan 1, NL–2288 EK Rijswijk (NL).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: QUAX, Wilhelmus, J. [NL/NL]; Jan Van Galenlaan 8, NL–2253 VB Voorschoten (NL).

(74) Agent: GLAISTER, Debra, J.; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304–1013 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INCREASING PRODUCTION OF PROTEINS IN GRAM–POSITIVE MICROORGANISMS

(57) Abstract

The present invention relates to secretion in Gram–positive microorganisms. The present invention provides the nucleic acid and amino acid sequences for the *Bacillus subtilis* secretion factors SecDF. The present invention also provides improved methods for the secretion of heterologous or homologous proteins in gram–positive microorganisms.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INCREASING PRODUCTION OF PROTEINS IN GRAM-POSITIVE MICROORGANISMS

FIELD OF THE INVENTION

The present invention generally relates to expression of proteins in gram-positive microorganisms and specifically to the gram positive microorganism secretion factor SecDF. The present invention provides expression vectors, methods and systems for the production of proteins in gram-positive microorganisms.

BACKGROUND OF THE INVENTION

Gram-positive microorganisms, such as members of the group *Bacillus*, have been used for large-scale industrial fermentation due, in part, to their ability to secrete their fermentation products into the culture media. In gram-positive bacteria, secreted proteins are exported across a cell membrane and a cell wall, and then are subsequently released into the external media usually obtaining their native conformation.

Secretion factors from Gram-positive microorganisms which have been identified and reported in the literature include SecA (Sadaie Y., Takamatsu h., Nakamura k., Yamane k.; Gene 98:101-105, 1991)., SecY (Suh J.-W., Boylan S.A., Thomas S.M., Dolan K.M., Oliver D.B., Price C.W.; Mol. Microbiol. 4:305-314, 1990)., SecE (Jeong S., Yoshikawa H., Takahashi H.; Mol. Microbiol. 10:133-142, 1993), FtsY an FfH (PCT/NL 96/00278), and PrsA (WO 94/19471).

By contrast, in the gram-negative microorganism, E.coli, protein is transported to the periplasm rather than across the cell membrane and cell wall and into the culture media. E.coli has at least two types of components of the secretory mechanism, soluble cytoplasmic proteins and membrane associated proteins. Reported E.coli secretion factors include the soluble cytoplasmic proteins, SecB and heat shock proteins; the peripheral membrane-associated protein SecA; and the integral membrane proteins SecY, SecE, SecD and SecF.

In spite of advances in understanding portions of the protein secretion machinery in procaryotic cells, the complete mechanism of protein secretion, especially for gram-positive microorganisms, such as *Bacillus*, has yet to be fully elucidated.

SUMMARY OF THE INVENTION

The capacity of the secretion machinery of a Gram-positive microorganism may become a limiting factor or bottleneck to protein secretion and the production of proteins in secreted form, in particular when the proteins are recombinantly introduced and overexpressed by the host cell. The present invention provides a means for alleviating that bottle neck.

The present invention is based, in part, upon the identification of the *Bacillus* secretion factor SecDF and upon the unexpected finding that, in contrast to SecD and SecF of *E.coli*, *Bacillus* SecDF is encoded by one nucleic acid sequence. The present invention is

also based upon the unexpected finding that SecDF has sequence as well as structural similarity to secondary solute transporters.

The present invention is also based, in part, upon the finding that SecDF mutants of *B.subtilis* have a cold-sensitive phenotype for growth and further that the rate of processing of exo-enzymes, amylase and neutral protease, is decreased in SecDF mutants of *B.subtilis*. The present invention is also based, in part, upon the finding that B.subtilis SecDF, which has twelve putative transmembrane domains is required for efficient translocation of secretory pre-proteins under conditions of hyper-secretion.

The present invention provides isolated nucleic acid and amino acid sequences for *B. subtilis* SecD, SecF and SecDF. The amino acid sequence and nucleic acid sequence for *B. subtilis* SecDF is shown in Figures 1A-1E.

The present invention also provides improved methods for secreting proteins from gram-positive microorganisms. Accordingly, the present invention provides an improved method for secreting desired proteins in a gram-positive microorganism comprising the steps of obtaining a gram positive microorganism comprising nucleic acid encoding at least one Bacillus secretion factor selected from the group consisting of SecD, SecF and SecDF wherein said secretion factor is under the control of expression signals capable of expressing said secretion factor in a gram-positive microorganism said microorganism further comprising nucleic acid encoding said protein; and culturing said microorganism under conditions suitable for expression of said secretion factor and secretion of said protein. In one embodiment of the present invention, the protein is homologous or naturally occurring in the gram-positive microorganism. In another embodiment of the present invention, the protein is heterologous to the gram-positive microorganism.

The present invention provides expression vectors and host cells comprising at least one nucleic acid encoding a gram-positive secretion factor selected from the group consisting of SecD, SecF and SecDF. In one embodiment of the present invention, the host cell is genetically engineered to produce a desired protein, such as an enzyme, growth factor or hormone. In yet another embodiment of the present invention, the enzyme is selected from the group consisting of proteases, carbohydrases including amylases, cellulases, xylanases, reductases and lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; transferases, kinases and phophatases acylases, amidases, esterases, oxidases.

In a further embodiment the expression of the secretion factor SecD, SecF and/or SecDF is coordinated with the expression of other components of the secretion machinery. Preferably other components of the secretion machinary, i.e., translocase, SecA, SecY, SecE and/or other secretion factors known to those of skill in the art are modulated in expression at an optimal ratio to SecD, SecF or SecDF. For example, it may be desired to overexpress multiple secretion factors in addition to SecDF for optimum enhancement of the secretion machinary.

The present invention also provides a method of identifying homologous non *Bacillus subtilis* secretion factors that comprises hybridizing part or all of *secDF* nucleic acid shown in Figures 1A-1E with nucleic acid derived from gram-positive microorganisms. In one embodiment, the nucleic acid is of genomic origin. In another embodiment, the nucleic acid is a cDNA. The present invention encompasses novel gram-positive secretion factors identified by this method.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1E shows the nucleic acid sequence for *secDF* (SEQ ID NO:1) and the deduced amino acid sequence of SecDF (SEQ ID NO:2).

Figure 2 shows the decreased rate of processing of pre-AmyQ in SecDF mutants of *B.subtilis*. Mutant strain (MIF1) and wildtype *B. subtilis* (168) harboring a plasmid encoding AmyQ (pKTH10; Takkinen K., Pettersson R.F., Kalkkinen N., Palva I., Soderlund H., Kaariainen L. J. Biol. Chem. 258:1007-1013(1983).) were tested for precursor and mature amylase using western blot analysis:

lane 1 and 2: proteins secreted into the medium

lane 3 and 4: total cell proteins analyzed.

Figures 3A-3B shows the expression of *secDF* in *B. subtilis* grown in TY medium (3A) and minimal media (3B) as measured by ß-gal.

Figure 4A-4B show the levels of AmyQ (*Bacillus amyloliquefaciens* α-amylase) accumulated in *B.subtilis* MIF1 relative to wildtype as a measurement of the total amount of AmyQ (Figure 4A) and as a percentage of pre-AmyQ (Figure 4B). Data are derived from the gel analysis of figure 2.

Figure 5A and 5B show a pulse chase experiment of amylase made in wild type *B. subtilis* and *B. subtilis* MIF1 (insertional inactivation of SecDF). Figure 5A is a 10% SDS gel with lane 2, 3 and 4 illustrating the levels of protein seen at 1' , 2' and 10' in wild type *B.subtilis* and lanes 5, 6 and 7 illustrating the levels of protein seen at 1', 2' and 10' in *B. subtilis* MIF1. After pulse chase the cells were lysed and the proteins were selectively precipitated with anti-amylase antibodies. Figure 5B shows the percentage of AmyQ precursor at chase times 1', 2', 5' and 10' of wild type *B. subtilis* and *B. subtilis* MIF1.

Figures 6A-6C. Figure 6A illustrates a chromosomal organization of the B.subtilis secDF locus (adapted from the Subtilist database; http://www.pasteur.fr.Bio/SubtiList .html). Figure 6B illustrates the chromosomal organization of the E.coli secD locus (adapted from Pogliano, et al., 1994, J. Bacteriol. 176:804-814 and Reuter et al., 1991, J. Bacteriol.

173:2256-2264). Figure 6C is a comparison of the deduced amino acid sequences of SecDF of B.subtilis and SecD (SEQ ID NO:3) and SecF (SEQ ID NO:4) of E.coli. Identical amino acids (*), or conservative replacements (.) are marked. The conserved regions D1-D6 and F1-F4, which are present in all known SecD and SecF proteins/domains are marked with black, or open bars. Putative membrane-spanning domains (I-XII) are indicated in gray shading. The membrane-spanning domains of E.coli SecD and SecF were adapted from Pogliano et al., 1994, J. Bacteriol. 176:804-814 and GenBank sequence ID number 134401, respectively. The membrane spanning domains in SecDF of B.subtilis were predicted using algorithms described by Sipos and von Heijne (Sipos et al., 1993, Eur. J. Biochem 213:1333-1340). The point of truncation of the SecDF protein in B.subtilis is indicated with an arrow.

Figure 7 illustrates the growth at 15°C as a function of time as measured at OD600 for the strains *B.subtilis* 168 pGS1 (neutral protease expression plasmid), *B.subtilis* 168 pKTH10 (amylase expression plasmid), *B.subtilis* MIF1 pKTH10 and *B.subtilis* MIF1 pGS1.

Figure 8 illustrates the genomic map of the nucleic acid encoding *secDF* and surrounding nucleic acid.

Figures 9A-9B illustrates the restriction map of plasmids MID2 (MID2 and MID refer to the same plasmid) (9A) and MIF1 (MIF and MIF1 refer to the same plasmid) (9B) containing internal *secDF* fragments which have been interrupted.

Figure 10: Demonstration that SecDF is a single protein in *B. subtilis*. A fusion was made between the ORF encoding *B. subtilis* SecDF and a c-myc polypeptide. This fusion protein was detected in a Western blot using antibodies directed to c-myc. It can be seen that a 97 kDa protein is detected corresponding to the expected size for a SecDF/myc fusion.

lane 1: overnight culture of *E. coli* (in TY) with plasmid pX-DFmyc
lane 2: e overnight culture *B. subtilis* 168 DF-myc (in TY) without xylose induction
lane 3: same as 2 grown with xylose induction

Size markers have been given (in kDa). pX-DFmyc was obtained from Dr. W. Schumann: it is a vector that will replicate in E.coli and integrate in *Bacillus*. The *secDF* gene has been cloned with a myc-tag at the C-terminus of SecDF. The *secDF* gene is under the control of the inducible xylose promoter.

Figure 11: Impaired extracellular accumulation of AmyQ. Cells overexpressing amylase were grown under two different conditions: at 37 °C during 1 hour and at 15 °C during 16 hours. The amount of secreted amylase was determined with Western blot analysis.

lane 1: *B. subtilis* 168 (pKTH10); medium after 1 hour growth at 37 °C

lane 2 :*B. subtilis* MIF1 (pKTH10); medium after 1 hour growth at 37 °C.

lane 3: *B. subtilis* 168 (pKTH10); medium after 16 hours growth at 15 °C

lane 4: *B. subtilis* MIF1 (pKTH10); medium after 16 hours growth at 15 °C

The bands have been scanned and analyzed: after 1 hour at 37 C strain MIF1 secretes 72 % compared to wildtype; after 16 hours at 15 C MIF1 secretes only 20 % compared to wild type level.

Figure 12; Impaired secretion of neutral protease. Cells overexpressing neutral protease (from plasmid GS1) were grown under two different conditions: at 37 °C during 1 hour and at 15 °C during 16 hours. The amount of secreted neutral protease was determined with Western blot analysis.

lane 1: *B. subtilis* 168 (pGS1); medium after 1 hour growth at 37 °C

lane 2 :*B. subtilis* MIF1 (pGS1); medium after 1 hour growth at 37 °C.

lane 3: *B. subtilis* 168 (pGS1); medium after 16 hours growth at 15 °C

lane 4: *B. subtilis* MIF1 (pGS1); medium after 16 hours growth at 15 °C

The amounts of neutral protease have been quantified: after 1 hour at 37 °C: MIF1 secretes 47% NprE compared to wildtype; after 16 hours at 15 °C: MIF1 secretes 43% NprE compared to wildtype.

Figure 13 shows the amino acid alignment of E.coli SecD (SEQ ID NO:3) with Bacillus subtilis SecDF (SEQ ID NO:2).

Figure 14 shows the amino acid alignment of E.coli SecF (SEQ ID NO:4) with Bacillus subtilis SecDF (SEQ ID NO:2).

Figure 15 shows the putative membrane-spanning domains numbered I-XIII. The positions of the patterns of conserved residues (D1-D6 and F1-F4) are indicated in bold. The carboxyl-terminus of the truncated SecDF protein of B.subtilis MIF is marked with and arrow. N is the amino-terminus and C is the carboxyl-terminus.

Figure 16A-16C. Figure 16A shows a schematic presentation of the secDF locus of B.subtilis MID. By a single-crossover event (Campbell-type integration), the secDF promoter region was replaced with the *Pspac* promoter of the integrated plasmid pMutin2, which can be repressed by the product of the *lacI* gene. Simultaneously, the *spoVG-lacZ* reporter gene of pMutin2 was placed under the transcriptional control of the *secDF* promoter region. The chromosomal fragment from the *secDF* regions which was amplified by PCR and cloned into pMutin2, is indicated with black bars. Only the restriction sites relevant for the construction are shown. *PsecDF* promoter region of the *secDF* gene; ori pBR322, replication functions of pBR322; secDF', 3' truncated *secDF* gene; T_1T_2, transcriptional terminators on pMutin2; SL, putative rho-independent terminator of *secDF* transcription. Figure 16B is a schematic

presentation of the *secDF* locus of B.subtilis MIF. The *secDF* gene was disrupted by the integrated plasmid pMutin2. *'secDF,5'* truncated *secDF* gene. Figure 16C shows the growth of secDF mutants in Ty medium at 15°C. Overnight cultures of strains grown in TY medium at 37°C were diluted 100-fold in fresh TY medium and incubated at 15 °C. Growth of B.subtilis 168 is shaded squares; 168 (pKTH10) open squares; MID closed triangle; MIDpKth10 open triangle; MIF closed circle; MIF (pKt1110) open circles in the absence of IPTG, was determined by optical density readings at 600nm. Growth of B.subtilis MID and MID (pKTH10) open triangles was determined in medium supplemented with IPTG.

Figures 17A-17D shows the identification of the SecDF protein in B.subtilis. To identify the SecDF protein, cells of B.subtilis XDF-Myc, which contain the *secDFmyc* gene under control of a xylose-inducible promoter, were grown in the absence or presence of xylose and protoplasted. In parallel, protoplasts were incubated for 30 min without further additions, in the presence of trypsin (1mg/ml), or in the presence of tyrpsin and Triton X-100 (1%). Samples were used for SDS-PAGE and Western blotting. Figure 17A illustrates that SecDF-Myc was visualized with specific antibodies against the c-Myc epitope. Figure 17B shows SipS (extracellular control) and Figure 17C shows GroEL (cytoplasmic control) which were visualized with specific antibodies. Figure 17D shows limited proteolysis of SecDF-Myc with trypsin (1mg/ml) that was performed by incubation for various periods of time. Intact SecDF-Myc (82kDa), trypsin resistant fragments of SecDF-Myc (54kDa and 23 kDa), SipS and GroEL are indicated.

DETAILED DESCRIPTION

Definitions

As used herein, the genus *Bacillus* includes all members known to those of skill in the art, including but not limited to *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus* and *B. thuringiensis*.

The present invention encompasses novel SecD, SecF and SecDF secretion factors from any gram positive organism. In a preferred embodiment, the gram-positive organism is *Bacillus*. In another preferred embodiment, the gram-positive organism is from *B. subtilis*. As used herein, the phrase, "*B.subtilis* SecDF secretion factor" refers to the amino acid sequence shown in Figures 1A-1E as well as the amino acid sequence encoded by the nucleic acid disclosed in Kunst et al., 1997, Nature 390:249-256 (GenBank accession number ID g2635229) and GenBank accession number AF024506 and the present invention encompasses the SecDF amino acid sequence encoded by secDF nucleic acid disclosed in Figures 1A-1E, GenBank accession number ID g2635229 and accession number AF024506. The present invention encompasses amino acid variants of Bacillus subtilis that are able to

modulate secretion alone or in combination with other secretions factors in gram-positive microorganisms.

As used herein, "nucleic acid" refers to a nucleotide or polynucleotide sequence, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be double-stranded or single-stranded, whether representing the sense or antisense strand. As used herein "amino acid" refers to peptide or protein sequences or portions thereof. As used herein, lower case "*secDF*" is used to designate a nucleic acid sequence, whereas upper case "SecDF" is used to designate an amino acid sequence. A "*B.subtilis* polynucleotide homolog" or "polynucleotide homolog" as used herein refers to a polynucleotide that has at least 80%, at least 90% and at least 95% identity to Figures 1A-1E or which is capable of hybridizing to part or all of the nucleic acid of Figures 1A-1E under conditions of high stringency and which encodes an amino acid sequence that is able to modulate secretion of the gram-positive microorganism from which it is derived. Modulate as used herein refers to the ability of a secretion factor to alter the secretion machinery such that secretion of proteins is altered.

The terms "isolated" or "purified" as used herein refer to a nucleic acid or amino acid that is removed from at least one component with which it is naturally associated.

As used herein, the term "heterologous protein" refers to a protein or polypeptide that does not naturally occur in a gram-positive host cell. Examples of heterologous proteins include enzymes such as hydrolases including proteases, cellulases, amylases, other carbohydrases, and lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; transferases, kinases and phophatases. The heterologous gene may encode therapeutically significant proteins or peptides, such as growth factors, cytokines, ligands, receptors and inhibitors, as well as vaccines and antibodies. The gene may encode commercially important industrial proteins or peptides, such as proteases, carbohydrases such as amylases and glucoamylases, cellulases, oxidases and lipases. The gene of interest may be a naturally occurring gene, a mutated gene or a synthetic gene.

The term "homologous protein" refers to a protein or polypeptide native or naturally occurring in a gram-positive host cell. The invention includes host cells producing the homologous protein via recombinant DNA technology. The present invention encompasses a gram-positive host cell having a deletion or interruption of the nucleic acid encoding the naturally occurring homologous protein, such as a protease, and having nucleic acid encoding the homologous protein, or a variant thereof, re-introduced in a recombinant form. In another embodiment, the host cell produces the homologous protein.

Detailed Description of the Preferred Embodiments

The present invention provides novel secretion factors and methods that can be used in gram-positive microorganisms to ameliorate the bottleneck to protein secretion and the production of proteins in secreted form, in particular when the proteins are recombinantly introduced and overexpressed by the host cell. The present invention provides the secretion factor SecDF derived from *Bacillus subtilis* and illustrates that interruption of the nucleic acid encoding SecDF via homologous recombination results in a loss in the host cell's capacity to process and secrete a recombinantly introduced heterologous pro-protein.

I. SecDF Nucleic Acid and Amino Acid Sequences

Nucleic Acid Sequences

The *secDF* polynucleotide having the sequence as shown in Figures 1A-1E and in Kunst et al., 1997, Nature 390:249-256 (GenBank accession number ID g2635229) encodes the *Bacillus subtilis* secretion factor SecDF. The *Bacillus subtilis* SecDF was initially identified via a FASTA search of *Bacillus subtilis* translated genomic sequences. The SecD and SecF portions of SecDF of Figures 1A-1E (see also Figures 13 and 14) were found to have 29 % and 28 % identity to *E.coli* SecD and SecF, respectively. Subsequent to Kunst et al., the B.subtilis nucleic acid sequence was confirmed and has been submitted to GenBank database with accession number AF024506. The present invention encompasses secDF nucleic acid disclosed in Figures 1A-1E, GenBank accession number ID g2635229 and accession number AF024506.

The present invention provides *secD* polynucleotide, *secF* polynucleotide and *secDF* polynucleotide which may be used alone or together in a host cell. The polynucleotide sequences for SecD and SecF portions of SecDF can be determined from Figures 13 and 14 which show the amino acid alignment of E. coli SecD and SecF with the Bacillus subtilis SecDF.

In contrast to *E.coli* secretion factors SecD and SecF and as illustrated in Figure 6, *Bacillus subtilis* SecDF is encoded by one polynucleotide. The SecD operon of E.coli consists of the YahC, secD and secF genes (Pogliano et al., 1994, J. Bacteriol. 176:804-814). This function-related operon structure is not conserved in B.subtilis, as the yajC-like gene yrbF and secDF are separated by two pairs of divergently transcribed genes, denoted yrzE, yrbG, spoVB and yrzD.

The present invention encompasses *secD*, *secF* and *secDF* polynucleotide homologs encoding gram-positive secretion factors SecD, SecF and SecDF, respectively, whether encoded by one or multiple polynucleotides which have at least 80%, or at least 90% or at least 95% identity to *B. subtilis* SecD, SecF and SecDF, respectively as long as the homolog encodes a protein that is able to function by modulating secretion in a gram-positive microorganism. As will be understood by the skilled artisan, due to the degeneracy of the genetic code, a variety of polynucleotides, i.e., *secD*, *secF* and *secDF* polynucleotide

variants, can encode the *Bacillus subtilis* secretion factors SecD, SecF and SecDF. The present invention encompasses all such polynucleotides.

Gram-positive microorganism polynucleotide homologs of *B. subtilis secD*, *secF* and *secDF* secretion factors can be identified through nucleic acid hybridization of gram-positive microorganism nucleic acid of either genomic of cDNA origin. The polynucleotide homolog sequence can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments disclosed in Figures 1A-1E. Accordingly, the present invention provides a method for the detection of *secD*, *secF* and *secDF* polynucleotide homologs which comprises hybridizing a nucleic acid sample with part or all of a nucleic acid sequence from *secD*, *secF* or *secDF*.

Also included within the scope of the present invention are *secDF*, *secD* and *secF* polynucleotide sequences that are capable of hybridizing to part or all of the *secDF* nucleotide sequence of Figures 1A-1E under conditions of intermediate to maximal stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA) incorporated herein by reference, and confer a defined "stringency" as explained below.

"Maximum stringency" typically occurs at about Tm-5°C (5°C below the Tm of the probe); "high stringency" at about 5°C to 10°C below Tm; "intermediate stringency" at about 10°C to 20°C below Tm; and "low stringency" at about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a maximum stringency hybridization can be used to identify or detect identical polynucleotide sequences while an intermediate or low stringency hybridization can be used to identify or detect polynucleotide sequence homologs.

The term "hybridization" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY).

The process of amplification as carried out in polymerase chain reaction (PCR) technologies is described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY). A nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides from the *secDF* nucleotide sequence of Figures 1A-1E, preferably about 12 to 30 nucleotides, and more preferably about 20-25 nucleotides can be used as a probe or PCR primer.

Amino Acid Sequences

The *B. subtilis secDF* polynucleotide as shown in Figures 1A-1E encodes *B. subtilis* SecDF. The B.subtilis secDF gene specifies one protein of 737 residues with a calculated molecular mass of 81,653. The SecDF protein has a two-domain structure, consisting of an amino-terminal SecD domain (about 416 residues) and a carboxyl-terminal SecF domain (291 residues). These domains show significant sequence similarity to known SecD and

SecF proteins from other organisms, the highest similarity being found with SecD and SecF proteins from the cyanobacterium Synechocystis. The stretch of 30 residues which links the SecD and SecF domains of B.subtilis SecDF is not conserved in other known SecD or SecF proteins. The corresponding domains of SecDF also show sequence similarity among themselves, in particular at their carboxyl-termini (22% identical residues and 44% conservative replacements in a stretch of 200 residues). B.subtilis SecDF shows amino acid sequence similarity to solute transporters, such as AcrF of E.coli (42% identical residues and conservative replacements in a stretch of 253 residues) which is involved in acriflavine resistance (GenBank sequence ID number g399429) and ActII-3 of Streptomyces coelicolor (46% identical residues and conservative replacements in a stretch of 159 residues) which is involved in the transport of antibiotics (GenBank sequence ID number g80715).

Alignment of B.subtilis SecDF with the SecD and SecF proteins from the organisms listed in Table I revealed that these proteins do not show similarity over their entire length. Ten short patterns of conserved amino acids were identified, which are present in all known SecD and SecF proteins. As shown in Figure 6C, these conserved regions were named D1-D6 and F1-F4 for the SecD and SecF domains/proteins, respectively. The positions of these conserved regions are indicated in Figure 6C. Some of these conserved domains are present in both SecD and SecF. This similarity is most obvious for the regions D1 and F1 which, respectively, have the consensus sequence G(L/I)DLRGG and G(L/I)DF(A/T)GG. Parts of the conserved regions D5 and F2 also show similarity.

The present invention encompasses gram positive microorganism amino acid variants of the amino acid sequence shown in Figures 1A-1E that are at least 80% identical, at least 90% identical and at least 95% identical to the sequence shown in Figures 1A-1E as long as the amino acid sequence variant is able to function by modulating secretion of proteins in gram-positive microorganisms.

Table I.
Percentage of identical residues plus conservative replacements in SecD and SecF domains and proteins from various organisms.

Organism	SecD	SecF
B. subtilis	100	100
E. coli	47	51
H. influenzae	48	52
H. pylori	45	49
M. jannaschii	39	39
M. tuberculosis	45	52
R. capsulatus	47	50
S. coelicolor	42	57
Synechocystis sp.	49	56

The GenBank sequence ID numbers are: SecD (*E. coli*) 134399; SecF (*E. coli*) 134401; SecD (*Huemophilus influenzae*) 1173414; SecF (*H. influenzae*) 1173416; SecD (*Helicobacter pylori*) 2314730; SecF (*H. pylori*) 2314729; SecD (*Methanococus jannaschii*) 2129225: SecF (*M. jannaschii*) 2129224; SecD (*Mycobacterium tuberculosis*) 2498898; SecF (*M. tuberculosis*) 2498900; SecD (*Rhodobacter capsulatus*) 2252773; SecF (*R. capsulatus*) 2252774; SecD (*S. coelicolor*) 1076081; SecF (*S. coelicolor*) 1076082; SecD (*Synechocystis sp.*) 1001493; SecF (*Synechocystis sp.*) 1001494.

II. Expression Systems

The present invention provides expression systems for the enhanced production and secretion of desired heterologous or homologous proteins in gram-positive microorganisms.

a. Coding Sequences

In the present invention, the vector comprises at least one copy of nucleic acid encoding a gram-positive microorganism SecD, SecF, or SecDF secretion factor and preferably comprises multiple copies. In a preferred embodiment, the gram-positive microorganism is *Bacillus*. In another preferred embodiment, the gram-positive microorganism is *Bacillus subtilis*. In a preferred embodiment, polynucleotides which encode *B. subtilis* SecD, SecD and/or SecDF, or fragments thereof, or fusion proteins or polynucleotide homolog sequences that encode amino acid variants of SecD, SecF and/or SecDF, may be used to generate recombinant DNA molecules that direct the expression of SecD, SecF, SecDF, or amino acid variants thereof, respectively, in gram-positive host cells. In a preferred embodiment, the host cell belongs to the genus *Bacillus*. In another preferred embodiment, the host cell is *B.subtilis*.

As will be understood by those of skill in the art, it may be advantageous to produce polynucleotide sequences possessing non-naturally occurring codons. Codons preferred by a particular gram-positive host cell (Murray E et al (1989) Nuc Acids Res 17:477-508) can be selected, for example, to increase the rate of expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

Altered gram positive *secD*, *secF* or *secDF* polynucleotide sequences which may be used in accordance with the invention include deletions, insertions or substitutions of different nucleotide residues resulting in a polynucleotide that encodes the same or a functionally equivalent SecD, SecF or SecDF homolog, respectively. As used herein a "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

As used herein an "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring gram positive *secD*, *secF* or *secDF*.

As used herein "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

The encoded protein may also show deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent gram-positive SecD, SecF or SecDF variant. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the variant retains the ability to modulate secretion. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine, phenylalanine, and tyrosine.

The *secD*, *secF* or *secDF* polynucleotides of the present invention may be engineered in order to modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, eg, site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns or to change codon preference, for example.

In one embodiment of the present invention, a *secDF*, *secD* or *secF* polynucleotide may be ligated to a heterologous sequence to encode a fusion protein. A fusion protein may also be engineered to contain a cleavage site located between the *secDF* nucleotide sequence and the heterologous protein sequence, so that the SecDF protein may be cleaved and purified away from the heterologous moiety.

b. Vector Sequences

Expression vectors used in expressing the secretion factors of the present invention in gram-positive microorganisms comprise at least one promoter associated with a secretion factor selected from the group consisting of SecD, SecF and SecDF, which promoter is functional in the host cell. In one embodiment of the present invention, the promoter is the wild-type promoter for the selected secretion factor and in another embodiment of the present invention, the promoter is heterologous to the secretion factor, but still functional in the host cell.

Additional promoters associated with heterologous nucleic acid encoding desired proteins or polypeptides may be introduced via recombinant DNA techniques. In one embodiment of the present invention, the host cell is capable of overexpressing a heterologous protein or polypeptide and nucleic acid encoding one or more secretion

factor(s) is(are) recombinantly introduced. In one preferred embodiment of the present invention, nucleic acid encoding the secretion factor is stably integrated into the microorganism genome. In another embodiment, the host cell is engineered to overexpress a secretion factor of the present invention and nucleic acid encoding the heterologous protein or polypeptide is introduced via recombinant DNA techniques. The present invention encompasses gram-positive host cells that are capable of overexpressing other secretion factors known to those of skill in the art, including but not limited to, SecA, SecY, SecE or other secretion factors known to those of skill in the art or identified in the future.

In a preferred embodiment, the expression vector contains a multiple cloning site cassette which preferably comprises at least one restriction endonuclease site unique to the vector, to facilitate ease of nucleic acid manipulation. In a preferred embodiment, the vector also comprises one or more selectable markers. As used herein, the term selectable marker refers to a gene capable of expression in the gram-positive host which allows for ease of selection of those hosts containing the vector. Examples of such selectable markers include but are not limited to antibiotics, such as, erythromycin, actinomycin, chloramphenicol and tetracycline.

c. Transformation

In one embodiment of the present invention, nucleic acid encoding one or more gram-positive secretion factor(s) of the present invention is introduced into a gram-positive host cell via an expression vector capable of replicating within the host cell. Suitable replicating plasmids for *Bacillus* are described in Molecular Biological Methods for *Bacillus*, Ed. Harwood and Cutting, John Wiley & Sons, 1990, hereby expressly incorporated by reference; see chapter 3 on plasmids. Suitable replicating plasmids for *B. subtilis* are listed on page 92.

In another embodiment, nucleic acid encoding one or more gram positive secretion factor(s) of the present invention are stably integrated into the microorganism genome. Preferred gram-positive host cells are from the genus *Bacillus*. Another preferred gram-positive host cell is *B. subtilis*. Several strategies have been described in the literature for the direct cloning of DNA in *Bacillus*. Plasmid marker rescue transformation involves the uptake of a donor plasmid by competent cells carrying a partially homologous resident plasmid (Contente *et al.*, *Plasmid* 2:555-571 (1979); Haima *et al.*, *Mol. Gen. Genet. 223*:185-191 (1990); Weinrauch *et al.*, *J. Bacteriol. 154(3)*:1077-1087 (1983); and Weinrauch *et al.*, *J. Bacteriol. 169(3)*:1205-1211 (1987)). The incoming donor plasmid recombines with the homologous region of the resident "helper" plasmid in a process that mimics chromosomal transformation.

Transformation by protoplast transformation is described for *B. subtilis* in Chang and Cohen, (1979) Mol. Gen. Genet 168:111-115; for B.megaterium in Vorobjeva et al., (1980)

FEMS Microbiol. Letters 7:261-263; for B. amyloliquefaciens in Smith et al., (1986) Appl. and Env. Microbiol. 51:634; for B.thuringiensis in Fisher et al., (1981) Arch. Microbiol. 139:213-217; for B.sphaericus in McDonald (1984) J. Gen. Microbiol. 130:203; and B.larvae in Bakhiet et al., (1985) 49:577. Mann et al., (1986, Current Microbiol. 13:131-135) report on transformation of *Bacillus* protoplasts and Holubova, (1985) Folia Microbiol. 30:97) disclose methods for introducing DNA into protoplasts using DNA containing liposomes.

III. Identification of Transformants

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the nucleic acid encoding a secretion factor is inserted within a marker gene sequence, recombinant cells containing the insert can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with nucleic acid encoding the secretion factor under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the secretion factor as well.

Alternatively, host cells which contain the coding sequence for a secretion factor and express the protein may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane-based, solution-based, or chip-based technologies for the detection and/or quantification of the nucleic acid or protein.

The presence of the *secDF* polynucleotide sequence can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments disclosed in Figures 1A-1E.

IV. Secretion Assays

Means for determining the levels of secretion of a heterologous or homologous protein in a gram-positive host cell and detecting secreted proteins include, using either polyclonal or monoclonal antibodies specific for the protein. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). These and other assays are described, among other places, in Hampton R et al (1990, Serological Methods, a Laboratory Manual, APS Press, St Paul MN) and Maddox DE et al (1983, J Exp Med 158:1211).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting specific polynucleotide sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the nucleotide sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially

available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced as shown in US Patent No. 4,816,567 and incorporated herein by reference.

V. Purification of Proteins

Gram positive host cells transformed with polynucleotide sequences encoding heterologous or homologous protein may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant gram-positive host cell comprising a secretion factor of the present invention will be secreted into the culture media. Other recombinant constructions may join the heterologous or homologous polynucleotide sequences to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins (Kroll DJ et al (1993) DNA Cell Biol 12:441-53).

Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals (Porath J (1992) Protein Expr Purif 3:263-281), protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequence such as Factor XA or enterokinase (Invitrogen, San Diego CA) between the purification domain and the heterologous protein can be used to facilitate purification.

Example I

Example I gives materials and methods for the Examples.

a. Plasmids, bacterial strains and media

Table II lists the plasmids and bacterial strains used herein. TY medium contained Bacto tryptone (1%), Bacto yeast extract (0.5%) and NaC1 (1%). S7 media 1 and 3, for the pulse-labeling of *B. subtilis* were prepared as described in van Diji et al. (1991, J. Gen. Microbiol. 137:2073-2083) with the exception that glucose was replaced by maltose. Minimal medium (GCHE medium; Kunst et al (1995, J. Bacteriol. 177: 2403-2407) contained glucose (1%), potassium L-glutamate (0.2%), potassium phosphate buffer (100 mM; pH 7), trisodium citrate (3 mM), $MgSO_4$ (3 mM), ferric ammonium citrate (22 mg/1), casein hydrolysate (0.1%), and L-tryptophan (50 mg/l). Antibiotics were used in the following concentrations:

chloramphenicol, 5 μg/ml; erythromycin, 1 μg/ml; kanamycin, 10 μg/ml; ampicillin, 50 μg/ml. IPTG was used at 1 mM.

b. DNA techniques

Procedures for DNA purification, restriction, ligation, agarose gel electrophoresis and transformation of competent *E. coli* DH5α cells were carried out as described in Sambrook (1989, Molecular cloning: a Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA). Enzymes were from Boehringer (Mannheim, Germany). *B. subtilis* was transformed by adding DNA to cells growing in GCHE medium at the end of the exponential growth phase, and continued incubation for 3-4 hours. PCR was carried out with Vent DNA polymerase (New England Biolabs, Beverly, MA), using buffers of the supplier. The nucleotide sequences of primers used for PCR (5'-3') are listed below; nucleotides identical to genomic template DNA are printed in capital letters and restriction sites used for cloning are underlined. DNA sequences were determined using the didioxy chain-termination procedure (Laemmli 1970, Nature, 227:680-685).

To verify the previously reported sequence of the *B. subtilis secF* gene (Kunst 1997 *supra*), a plasmid (pSecDF) was constructed by inserting a DNA fragment containing the entire *secDF* gene, amplified by PCR with the primers AB34secd (aaaagcttAAGGGAGGATATACATAATG) and AB37secd (aaggatccGCGTATGTCATTATAGC), into the *Hind*III and *Bam*HI restriction sites of the phagemid pBluescript II+.

To construct *B. subtilis* MIF an internal fragment of the *secDF* gene (417 nucleotides) was amplified by PCR with the oligonucleotides AB32secF (aaaagcttCGACAGAGCAAGTTGAG) and AB33secF (aaggatccGATTGTATCGTTAATGG) and, subsequently, cloned into pMutin2, which resulted in plasmid pMIF. To construct *B. subtilis* MID a fragment containing the ribosome binding site, start codon and the first 879 nucleotides of the *secDF* gene, but not the *secDF* promoter(s), was amplified with the primers AB34secD (see above) and AB31secD (aaggatccGTGTAATGTAGATATAAAC) and cloned into pMutin2, resulting in plasmid pMID. *B. subtilis* MIF and MID were obtained by Campbell type integration of plasmids pMIF and PMID, respectively, into the chromosome of *B. subtilis* 168. Correct integration of plasmids in the chromosome of *B. subtilis* was verified by Southern hybridization. To construct *B. subtilis* XDF-Myc the entire secDF gene was amplified by PCR with the primers AB47secD (aatctagaAAGGGAGGATATACATAATG) and AB46mycF (a*ggatcc***ttagttcaaatcttcctcactgatcaatttctg**TTCTTGCGCCGAATCTTTTTTCAG); the sequence specifying the human c-Myc epitope is indicated in bold). The resulting PCR

product, which contains the *secDFmyc* gene, was cleaved with *Xba*I and *Bam*HI, and ligated into the *Spe*I and *Bam*HI sites of pX. This resulted in plasmid pXDFmyc, which contains the *secDFmyc* gene under the transcriptional control of the xylose-inducible *xylA* promoter. Upon transformation of *B. subtilis* 168 with pXDFmyc, both the *xylA* promoter and *secDFmyc* were integrated into the chromosomal *amyE* gene, resulting in *B. subtilis* XDF-Myc. The disruption of the *amyE* gene was confirmed by growing *B. subtilis* XDF-Myc on TY plates containing 1% starch and subsequent exposure of the plates to iodine. As shown by a lack of halo formation, *B. subtilis* XDF-Myc did not secrete active α-amylase.

c. Pulse-chase protein labeling, immunoprecipitation, *SDS-PAGE and fluorography*- Pulse-chase labeling experiments with *B. subtilis* and immunoprecipitations were performed as described in van Dijl et al., 1991, J. Gen. Microbiol 137:2073-2083. SDS-PAGE was performed according to Laemmli (1970, Nature 227:680-685). [^{14}C]-methylated molecular weight markers were from Amersham (Little Chalfont, UK). Fluorography was performed with Autofluor (National Diagnostics, Atlanta, Georgia, USA). Relative amounts of precursor and mature forms of secreted proteins were estimated by scanning of autoradiographs with an LKB ultrascan XL laser densitometer (LKB, Bromma, Sweden).

d. Western blot analysis- Western blotting was performed using a semi-dry system as described in Miller supra. After separation by SDS-PAGE, proteins were transferred to Immobilon-PVDF membranes (Millipore Corp., Bedford, MA). Proteins were visualized with specific antibodies and horseradish peroxidase (HRP) anti-rabbit or anti-mouse IgG conjugates, using the ECL detection system of Amersham. Streptavidin-HRP conjugate was obtained from Amersham.

e. Protease accessibility- Protoplasts were prepared from exponentially growing cells of *B. subtilis*. To this purpose cells were concentrated 5-fold in protoplast buffer (20 mM potassium phosphate, pH 7.5; 15 mM $MgCl_2$; 20% sucrose) and incubated for 30 min in the presence of 1 mg/ml lysozyme (37°C). Next, the protoplasts were collected by centrifugation and resuspended in fresh protoplast buffer. The protease accessibility of membrane proteins was tested by incubating the protoplasts at 37°C in the presence of 1 mg/ml trypsin (Sigma Chemical Co., St. Louis, MO, USA) for various periods of time. The reaction was terminated by the addition of 1.2 mg/ml trypsin inhibitor (Sigma Chemical Co.). Finally, protoplasts were collected by centrifugation, and the degradation of specific proteins was analyzed by SDS-PAGE and Western-blotting. In parallel, protoplasts were incubated without trypsin, or in the

presence of trypsin and 1% Triton X-100. Samples containing TX-100 were directly used for SDS-PAGE after the addition of trypsin inhibitor.

f. β-Galactosidase activity- Overnight cultures were diluted 100-fold in fresh medium and samples were taken at hourly intervals for optical density (OD) readings at 600 nm and β-Galactosidase activity determinations. The β-Galactosidase assay and the calculation of β-Galactosidase units (per OD600) were performed as described in Miller, 1982, Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Table II
Plasmids and Bacterial Strains

Strain/Plasmid	Genotype/Properties	Source/Reference
Strains		
E. coli		
DH5α	*F80dlacZ-M15 endAI recAI hsdR17(r_k. m_{k+}) thi-I gyrA96 relAI-(lacZYA-argF)* U169	Bethesda Research Laboratories
B. subtilis		
168	*trpC2*	Kunst et al (supra)
MIF	*trpC2; secDF*::pMIF; Em^r	Examples
MID	*trpC2; secDF*::pMID; Em^r	Examples
XDF-Myc	*trpC2; amyE::xylA-secDFmyc*;Cm^r	Examples
Plasmids		
pBluescript II KS+	cloning vector; Ap^r	Stratagene
pSecDF	pBluescript II KS+ derivative; carries the *B. subtilis secDF* gene	Examples
pX	vector for the integration of genes in the *amyE* locus of *B. subtilis;* integrated genes will be transcribed from the *xylA* promoter; carries the *xylR* gene; Ap^r; Cm^r	Kim et al. 1996, Gene 181: 71-76
pXDFmyc	pX derivative; carries the *B. subtilis secDFmyc* gene downstream of the *xylA* promoter	Examples
pMutin2	pBR322-based integration vector for *B. subtilis;* contains a multiple cloning site downstream of the *Pspac* promoter (Yansura et al., 1984, Genetics and Biochemistry of Bacilli pp. 249-263 Academic Press, Orlando, USA), and a promoterless *lacZ* gene preceded by the ribosome-binding site of the *spoVG* gene; Ap^r; Em^r	V. Vagner and S.D. Ehrlich
pMIF	pMutin2 derivative; carries an internal fragment of the *secDF* gene	Examples
pMID	pmutin2 derivative; carries the 5' part of the *B. subtilis secDF* gene	Examples
pKTH10	Contains the *amyQ* gene of *B. amyloliquefaciens;* Km^r	Palva, 1982, Gene 19:81-87
pKTH10-BT	pKTH10 derivative, encodes the AmyQ-PSBT fusion protein	Tjalsma et al. 1998

Example II

This Example describes the membrane topology of Bacillus subtilis SecDF. Algorithms described by Sipos and von Heijne (Sipos et al.,1993, Eur. J. Biochem 213:1333-1340) predict that the *SecDF* (Bsu) protein has twelve membrane-spanning domains, the amino- and carboxyl-termini being localized in the cytoplasm. Two large extracellular loops are localized between the first and second, and the seventh and eighth membrane-spanning domains, respectively (Fig. 15). These predictions are in good agreement with the topology models proposed for SecD and SecF of *E. coli,* in which both SecD and SecF have six membrane-spanning domains with large periplasmic loops being located between the first and second membrane-spanning domains (Pogliano, 1994, *supra*).

To verify the predicted cytoplasmic localization of the carboxyl-terminus of *SecDF*, we studied the protease-accessibility of *SecDF*-Myc in protoplasts. As shown by Western blotting, two trypsin-resistant *SecDF*-Myc-derived fragments of about 54 kDa and 23 kDa were detectable upon incubation of intact protoplasts of xylose-induced *B. subtilis* XDF-Myc cells with trypsin. Under the same conditions, the *B. subtilis* signal peptidase SipS, of which a large part is exposed to the external side of the membrane (van Dijl et al, EMBO J. 11:2819-2828), was completely degraded by trypsin, whereas the cytoplasmic protein GroEL remained unaffected. In contrast, both *SecDF*-Myc-derived fragments and GroEL were completely degraded by trypsin when protoplasts were lysed by the addition of 1% Triton X-100. Taken together, these findings show that the carboxyl-terminus of *SecDF*-Myc is protected against trypsin in intact protoplasts, suggesting that the carboxyl-terminus of B.subtilis *SecDF* is localized in the cytoplasm.

To study the kinetics of the formation of the two trypsin-resistant *SecDF*-Myc-derived fragments, limited proteolysis experiments were performed in which protoplasts of xylose-induced *B. subtilis* XDF-Myc cells were incubated with trypsin for various periods of time. As shown by Western blotting, the 54 kDa fragment is a transiently existing intermediate product in the degradation of intact *SecDF*-Myc to the trypsin-resistant 23 kDa fragment. As judged from the apparent molecular masses of the trypsin-resistant fragments, it is most likely that trypsin cleavage of *SecDF*-Myc occurs in the two predicted extracellular domains between the first and second membrane-spanning domains, and the seventh and eighth membrane-spanning domains.

Example III

This Example relates to the cold-sensitive growth of B.subtilis secDF mutants. To analyze the effects of *SecDF* depletion on cell growth and protein secretion, two mutant *B. subtilis* strains were constructed with the integrative plasmid pMutin2 (provided by V. Vagner

and S.D. Ehrlich, INRA, Jouy en Josas, France). In the first strain, denoted *B. subtilis* MID, the encoding sequence of the *secDF* gene was left intact, but the *secDF* promoter was replaced with the IPTG-inducible *Pspac* promoter, present on pMutin2; in the second strain, denoted *B. subtilis* MIF, the coding sequence of the *SecDF* gene was disrupted with pMutin 2 (Fig. 16A and 16B, respectively). The point of truncation of the *SecDF* protein of *B. subtilis* MIF is indicated in Fig.6C. Irrespective of the growth medium used or the presence of IPTG, both *B. subtilis* MID and MIF showed growth rates at 37°C similar to that of the parental strain *B. subtilis* 168, demonstrating that under these conditions *SecDF* was not essential for growth and viability of the cells. By contrast, *SecDF* was important for growth in TY medium at 15°C: compared to the growth of the parental strain (Fig. 16C, indicated by the closed square), the growth of *B. subtilis* MID (in the absence of IPTG) and *B. subtilis* MIF was significantly reduced. In fact, the growth rates of the two latter strains were reduced to the same extent (Fig. 16C, indicated by the closed triangle and circle respectively) and, in addition, the cells of both strains showed a filamentous morphology. Growth of *B. subtilis* MID AT 15°C could be restored by the addition of IPTG to the growth medium (Fig. 16C, indicated with the closed triange), though not completely to wild-type levels. Similarly, growth of *B. subtilis* MIF at 15°C could be restored to a similar level as that of *B. subtilis* MID in the presence of IPTG, by introducing the *secDF-myc* gene in the *amyE* locus, indicating the c-Myc tag did not interfere with *SecDF* function. Interestingly, the growth defects of *B. subtilis* MID (in the absence of IPTG) and MIF were not observed instantaneously upon incubation at 15°C, as both strains showed growth rates comparable to those of the parental strain until the mid-exponential growth phase (OD600=0.3-0.4; Fig. 16C).

To test whether *SecDF* might be even more important for growth under conditions of hyper-secretion, the *B. subtilis* MID and MIF strains were transformed with plasmid pKTH10, which results in the secretion of the *Bacillus amyloliquefaciens* α-amylase AmyQ at high levels (≈1.3 g/l; Kontinen et al., 1988, J. Gen Microbiol, 134:2333-2344 and Palva, 1982, Gene 19:81-87). Irrespective of the presence of pKTH10, growth of *B. subtilis* MID and MIF at 37°C was not affected. In contrast, at 15 °C *B. subtilis* MID (in the absence of IPTG) and MIF cells transformed with pKTH10 completely stopped growing after reaching the mid-exponential growth phase and, subsequently, cells even started to lyse (Fig. 16C, indicated with the open triangle and circle, respectively). The latter observation showed that the cold-sensitive phenotype of cells depleted of *SecDF* was exacerbated by high levels of AmyQ secretion. The presence of pKTH10 did not affect the growth at 15°C of either the parental strain, or *B. subtilis* MID in the presence of IPTG (Fig. 16C, indicated with the open triangle), showing that high-level secretion of AmyQ *per se* did not affect the growth of *B. subtilis* at

low temperature. Taken together, these observations show that the B.subtilis *SecDF* (Bsu) protein is required for efficient growth at low temperatures, in particular under conditions of high-level protein secretion.

Example IV

This Example demonstrates that SecDF is required for efficient secretion of AmyQ. To investigate the importance of *SecDF* for protein secretion at moderate levels (about 30 mg of protein per liter), the secretion of the neutral protease NprE by *B. subtilis* MIF was analyzed by Western blotting. Both at 37°C and 15°C, the absence of *SecDF* did not result in the accumulation of pre-NprE, and similar amounts of mature NprE were detected in the medium of *B. subtilis* MIF and the parental strain.

To evaluate the importance of *SecDF* under conditions of hyper-secretion, the secretion of AmyQ into the growth medium was investigated by Western blotting experiments. The results showed that *B. subtilis* MIF (pKTH10) secreted reduced levels of AmyQ into the culture medium. This was most clearly observed with cells in the transition phase between exponential and post-exponential growth., which had been washed and resuspended in fresh medium. If the washed cells were incubated for 1 hour at 37°C, the medium of *B. subtilis* MIF contained about 65% ± 10%of the amount of AmyQ secreted by the parental strain. An even more drastic effect was observed at 15°C; after 16 hours of incubation, the medium of *B. subtilis* MIF contained about 40% ± 10% of the amount of AmyQ secreted by the parental strain. The reduced secretion of AmyQ into the medium by *B. subtilis* MIF was paralleled by an increased accumulation of pre-AmyQ in the cells. Since the cellular levels of mature AmyQ were not affected in the absence of intact *SecDF*, these data suggest that *SecDF* is required for the efficient translocation of pre-AmyQ, but not the release of mature AmyQ from the membrane.

To investigate the important of *SecDF* for the translocation of pre-AmyQ, *B. subtilis* MIF was transformed with plasmid pKTH10-BT^2, which specifies a hybrid AmyQ protein containing the biotin-accepting domain (PSBT) of a transcarboxylase from *Propionibacterium shermannii* (Jander et al., 1996, J. Bacteriol. 178:3049-3058) fused to its carboxyl-terminus. The rationale of this experiment is that pre-AmyQ-PBST will only be biotinylated by the cytoplasmic biotin-ligase if the rate of translocation of pre-AmyQ-PSBT is slowed down to such an extent that the PSBT-domain can fold into its native three-dimensional structure and accept biotin before transport across the membrane.

To investigate the importance of *SecDF* for the translocation of pre-AmyQ, *B. subtilis* MIF was transformed with plasmid pKTH10-BT^2, which specifies a hybrid AmyQ protein

containing the biotin-accepting domain (PSBT) of a transcarboxylase from *Propionibacterium shermannii* (Jander, *supra*) fused to its carboxyl-terminus. The rationale of this experiment is that pre-AmyQ-PSBT will only be biotinylated by the cytoplasmic biotin-ligase if the rate of translocation of pre-AmyQ-PSBT is slowed down to such an extent the PSBT-domain can fold into its native three-dimensional structure and accept biotin before transport across the membrane. Cells lacking intact *SecDF* accumulate biotinylated pre-AmyQ-BT, whereas no biotinylated (pre-)AmyQ-PSBT was detected in cells of the parental strain of *B. subtilis* XDF-Myc, which were transformed with pKTH10-BT. These finds show that the rate of translocation of pre-AmyQ-PSBT is significantly reduced in cells lacking *SecDF*.

To determine the rate of pre-AmyQ translocation in the absence of *SecDF*, the kinetics of pre-AmyQ processing by signal peptidase were studied by pulse-chase labeling of *B. subtilis* MIF containing pKTH10. Even at 37°C the rate of pre-AmyQ processing was decreased in cells lacking an intact *SecDF* gene; after a chase of 1 min, about 32% of the labeled AmyQ was mature in *B. subtilis* MIF whereas, under the same conditions, about 59% of the AmyQ was mature in the parental strain. The effects of the absence of intact *SecDF* were even more pronounced at 23°C; after a chase of 4 min, mature AmyQ was hardly detectable in *B. subtilis* MIF whereas, under the same conditions, about 40% of the labeled AmyQ was mature in the parental strain.

Pulse-chase labeling experiments were also performed with *B. subtilis* XDF-Myc, which overproduces the *SecDF*-Myc protein upon induction with xylose. Overproduction of *SecDF*-Myc did not significantly influence the rate of pre-AmyQ processing, showing that wild-type levels of *SecDF* are not limiting for the translocation of pre-AmyQ and that overproduction of *SecDF-myc* does not interfere with normal *SecDF* function.

Example V

This example describes the growth phase and medium-dependent transcription of the secDF gene. To test whether the transcription of the *secDF* gene depends on the growth phase or medium composition, as previously shown for the signal peptidase-encoding genes *sipS* and *sipT* (Bolhuis et al., 1996, Mol. Microbiol. 22:605-618 and Tjalsma, 1997, J. Biol. Chem., 272: 25983-25992), we made use of the transcriptional *secDF-lacZ* gene fusions present in *B. subtilis* MID and MIF. *B.subtilis* MIF was grown in three different media (minimal medium, TY, or TY supplemented with 1% glucose), and samples withdrawn at hourly intervals were assayed for β-galactosidase activity. Nearly constant levels of β-galactosidase activity were observed during growth in minimal medium, suggesting that the *secDF* gene was expressed constitutively. In contrast, cells grown in TY medium showed

increasing levels of β-galactosidase activity during exponential growth, with a maximum at the beginning of the stationary phase. The β-galactosidase activity decreased in the post-exponential growth phase suggesting that *secDF* promoter activity was highest in the transition phase between the exponential and post-exponential growth phase. The addition of 1% glucose to TY medium caused a drastic increase in the β-galactosidase levels of cells in the post-exponential growth phase, showing that glucose strongly stimulates the transcription of the *secDF* gene. Taken together, these findings show that the transcription levels of the *secDF* gene depend on the growth phase and growth medium

Example VI

This Example illustrates that secDF encodes one protein. To show that the *secDF* gene encodes only one protein of approximately 82 kDa, the 3' end of the *secDF* gene was extended with 11 codons, specifying the human c-*Myc* epitope (EQKLISEEDLN; Evan et al., 1985, Mol. Cell. Biol. 5: 3610-3616). Next, the *myc*-tagged *secDF* gene (*secDF*-myc) was placed under the transcriptional control of the xylose-inducible *xylA* promoter and, subsequently, integrated *via* a double-crossover replacement recombination into the *amyE* locus of *B. subtilis*, using the pX system developed by Kim et al. (1996, Gene 181:71-76). The resulting strain was named *B. subtilis* XDF-*Myc*. As shown by Western blotting and subsequent immuno-detection with c-*Myc*-specific monoclonal antibodies, the *SecDF*-Myc protein was produced in *B. subtilis* XDF-Myc cells growing in TY medium supplemented with 1% xylose, but not in cells growing in TY medium lacking xylose. Similar results were obtained if samples for Western blotting were prepared from intact cells or protoplasts of *B. subtilis* XDF-Myc. Immunodetection with SecDF-specific antibodies showed that the SecDF-Myc protein was highly overproduced in xylose-induced cells of *B. subtilis* XDF-Myc, as neither wild-type SecDF nor SecDF-Myc were detectable in uninduced cells. Judged from its mobility on SDS-PAGE, *SecDF-Myc* is a protein of about 82 kDa, which is in agreement with the sequence-based prediction.

Example VII

Detection of gram-positive microorganisms

The following example describes the detection of gram-positive microorganism SecDF.

DNA derived from a gram-positive microorganism is prepared according to the methods disclosed in Current Protocols in Molecular Biology, Chap. 2 or 3. The nucleic acid is subjected to hybridization and/or PCR amplification with a probe or primer derived from

SecDF. A preferred probe comprises the nucleic acid section containing conserved amino acid sequences.

The nucleic acid probe is labeled by combining 50 pmol of the nucleic acid and 250 mCi of [gamma ^{32}P] adenosine triphosphate (Amersham, Chicago IL) and T4 polynucleotide kinase (DuPont NEN®, Boston MA). The labeled probe is purified with Sephadex G-25 super fine resin column (Pharmacia). A portion containing 10^7 counts per minute of each is used in a typical membrane based hybridization analysis of nucleic acid sample of either genomic or cDNA origin.

The DNA sample which has been subjected to restriction endonuclease digestion is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40 degrees C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. The blots are exposed to film for several hours, the film developed and hybridization patterns are compared visually to detect polynucleotide homologs of *B.subtilis* SecDF. The homologs are subjected to confirmatory nucleic acid sequencing. Methods for nucleic acid sequencing are well known in the art. Conventional enzymatic methods employ DNA polymerase Klenow fragment, SEQUENASE® (US Biochemical Corp, Cleveland, OH) or Taq polymerase to extend DNA chains from an oligonucleotide primer annealed to the DNA template of interest.

Various other examples and modifications of the foregoing description and examples will be apparent to a person skilled in the art after reading the disclosure without departing from the spirit and scope of the invention, and it is intended that all such examples or modifications be included within the scope of the appended claims. All publications and patents referenced herein are hereby incorporated in their entirety.

CLAIMS

1. An expression vector comprising nucleic acid encoding at least one gram-positive microorganism secretion factor selected from the group consisting of SecD, SecF and SecDF, wherein said secretion factor is under the control of expression signals capable of overexpressing said secretion factor in a gram-positive organism.

2. The expression vector of Claim 1 wherein said gram-positive organism is a *Bacillus*.

3. The expression vector of Claim 1 wherein the *Bacillus* is selected from the group consisting of *B. subtilis*, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. circulans, B. lautus and *Bacillus* thuringiensis.

4. A gram positive microorganism comprising the expression vector of Claim 1.

5. The microorganism of Claim 4 that is a *Bacillus*.

6. The microorganism of Claim 5 wherein the *Bacillus* is selected from the group consisting of *B. subtilis*, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. circulans, B. lautus and *Bacillus* thuringiensis.

7. The microorganism of Claim 4 wherein said host cell is capable of expressing a heterologous protein.

8. The microorganism of Claim 7 wherein said heterologous protein is selected from the group consisting of hormone, enzyme, growth factor and cytokine.

9. The microorganism of Claim 8 wherein said heterologous protein is an enzyme.

10. The microorganism of Claim 9 wherein said enzyme is selected from the group consisting of a proteases, cellulases, amylases, carbohydrases, and lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; transferases, kinases and phophatases.

11. An improved method for secreting a protein in a gram-positive microorganism comprising the steps of obtaining a gram positive microorganism comprising nucleic acid encoding at least one Bacillus secretion factor selected from the group consisting of SecD, SecF and SecDF wherein said secretion factor is under the control of expression signals capable of expressing said secretion factor in a gram-positive microorganism said microorganism further comprising nucleic acid encoding said protein; and culturing said microorganism under conditions suitable for expression of said secretion factor and secretion of said protein.

12. The method of Claim 11 wherein said protein is homologous to said host cell.

13. The method of Claim 11 wherein said protein is heterologous to said host cell.

14. The method of Claim 11 wherein said gram-positive organism is a *Bacillus.*

15. The method of Claim 14 wherein said *Bacillus* is selected from the group consisting of *B. subtilis, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. circulans, B. lautus and Bacillus thuringiensis.*

16. The method of Claim 13 wherein said heterologous protein is selected from the group consisting of a hormones enzymes growth factor and cytokine.

17. The method of Claim 16 wherein said heterologous protein is an enzyme.

18. The method of Claim 17 wherein said enzyme is selected from the group consisting of a proteases, cellulases, amylases, carbohydrases, reductases and lipases, isomerases, transferases, kinases and phophatases.

```
          10                  30
ATGAAAAAAGGACGCTTGATTGCGTTTTTCCTTTTCGTTCTATTGATC
M  K  K  G  R  L  I  A  F  F  L  F  V  L  L  I

50                  70                  90
GGCACGGGCTTGGGCTACTTTACGAAGCCTGCCGCTAACAATATTACG
G  T  G  L  G  Y  F  T  K  P  A  A  N  N  I  T

           110                 130
TTAGGATTGGATTTGCAAGGCGGATTTGAGGTGCTGTATGATGTACAG
L  G  L  D  L  Q  G  G  F  E  V  L  Y  D  V  Q

   150                 170                 190
CCTGTAAAAAAAGGTGACAAAATCACAAAAGACGTTCTGGTCAGCACA
P  V  K  K  G  D  K  I  T  K  D  V  L  V  S  T

               210                 230
GTAGAGGCACTGAACCGCCGGGCCAATGTTCTCGGTGTCAGCGAACCG
V  E  A  L  N  R  R  A  N  V  L  G  V  S  E  P

       250                 270                 2
AACATCCAAATTGAAGGGAATAACCGGATTCGCGTTCAGCTCGCTGGC
N  I  Q  I  E  G  N  N  R  I  R  V  Q  L  A  G

90                 310                 330
GTGACAAACCAAAACAGAGCGCGTGAAATTTTGGCGACTGAAGCGCAG
V  T  N  Q  N  R  A  R  E  I  L  A  T  E  A  Q

           350                 370
CTTTCTTTCAGAGATGCAAATGATAAGGAACTGTTAAACGGTGCTGAT
L  S  F  R  D  A  N  D  K  E  L  L  N  G  A  D

   390                 410                 430
CTAGTCGAAAACGGCGCTAAACAAACTTATGATAGCACAACAAATGAG
L  V  E  N  G  A  K  Q  T  Y  D  S  T  T  N  E

               450                 470
CCAATTGTCACGATTAAGCTGAAAGACGCTGATAAATTTGGTGAAGTG
P  I  V  T  I  K  L  K  D  A  D  K  F  G  E  V
```

GC 385 Figure 1A

```
       490                   510                   5
ACCAAGAAGGTCATGAAAATGGCGCCAAACAACCAGCTTGTCATTTGG
T  K  K  V  M  K  M  A  P  N  N  Q  L  V  I  W

30                   550                   570
TTGGATTATGATAAAGGTGATTCCTTTAAGAAAGAAGTTCAAAAAGAG
L  D  Y  D  K  G  D  S  F  K  K  E  V  Q  K  E

           590                   610
CATCCTAAATTTGTATCCGCTCCAAATGTAAGTCAGGAACTAAATACA
H  P  K  F  V  S  A  P  N  V  S  Q  E  L  N  T

   630                   650                   670
ACTGATGTAAAAATTGAAGGTCATTTCACAGCTCAAGAAGCGAAAGAT
T  D  V  K  I  E  G  H  F  T  A  Q  E  A  K  D

                690                   710
TTAGCCAGCATTTTAAACGCAGGCGCACTTCCTGTGAAACTGACTGAA
L  A  S  I  L  N  A  G  A  L  P  V  K  L  T  E

       730                   750                   7
AAGTATTCGACATCAGTAGGCGCGCAATTCGGCCAGCAGGCTCTCCAT
K  Y  S  T  S  V  G  A  Q  F  G  Q  Q  A  L  H

70                   790                   810
GATACGGTGTTTGCCGGTATTGTCGGTATCGCAATTATTTTCTTATTT
D  T  V  F  A  G  I  V  G  I  A  I  I  F  L  F

           830                   850
ATGCTTTTCTATTACCGTCTGCCGGGATTAATCGCGGTGATTACGCTG
M  L  F  Y  Y  R  L  P  G  L  I  A  V  I  T  L

   870                   890                   910
TCTGTTTATATCTACATTACACTCCAGATCTTTGACTGGATGAATGCC
S  V  Y  I  Y  I  T  L  Q  I  F  D  W  M  N  A

                930                   950
GTACTCACGCTTCCGGGAATTGCCGCTCTCATTTTAGGTGTCGGGATG
V  L  T  L  P  G  I  A  A  L  I  L  G  V  G  M

       970                   990                  10
GCTGTTGACGCCAACATTATTACCTATGAGCGGATTAAAGAAGAGCTC
A  V  D  A  N  I  I  T  Y  E  R  I  K  E  E  L
```

GC 385 Figure 1B

```
10                  1030                  1050
AAGCTAGGAAAGTCAGTCCGCTCTGCCTTCCGTTCAGGAAACAGACGG
K  L  G  K  S  V  R  S  A  F  R  S  G  N  R  R

          1070                  1090
TCATTTGCGACGATTTTTGACGCGAATATTACAACCATTATTGCAGCG
S  F  A  T  I  F  D  A  N  I  T  T  I  I  A  A

  1110                  1130                  1150
GTTGTGCTCTTTATCTTTGGGACAAGCTCTGTTAAAGGGTTTGCGACA
V  V  L  F  I  F  G  T  S  S  V  K  G  F  A  T

               1170                  1190
ATGCTGATCCTATCGATTTTGACAAGCTTTATCACTGCCGTTTTCTTA
M  L  I  L  S  I  L  T  S  F  I  T  A  V  F  L

      1210                  1230                  12
TCGAGATTTCTCCTCGCTCTCCTTGTGGAAAGCAGATGGCTTGATCGG
S  R  F  L  L  A  L  L  V  E  S  R  W  L  D  R

50                  1270                  1290
AAAAAAGGCTGGTTTGGTGTCAATAAGAAACATATCATGGATATTCAG
K  K  G  W  F  G  V  N  K  K  H  I  M  D  I  Q

          1310                  1330
GATACGGATGAAAATACAGAGCCGCATACGCCATTCCAAAAATGGGAT
D  T  D  E  N  T  E  P  H  T  P  F  Q  K  W  D

  1350                  1370                  1390
TTCACGAGCAAACGCAAATACTTCTTTATTTTCTCCAGTGCGGTCACG
F  T  S  K  R  K  Y  F  F  I  F  S  S  A  V  T

               1410                  1430
GTTGCCGGGATTATTATCCTGCTTGTGTTCAGGCTGAATCTTGGCATT
V  A  G  I  I  I  L  L  V  F  R  L  N  L  G  I

      1450                  1470                  14
GACTTTGCAAGCGGTGCACGGATTGAAGTGCAAAGCGACCATAAGCTG
D  F  A  S  G  A  R  I  E  V  Q  S  D  H  K  L
```

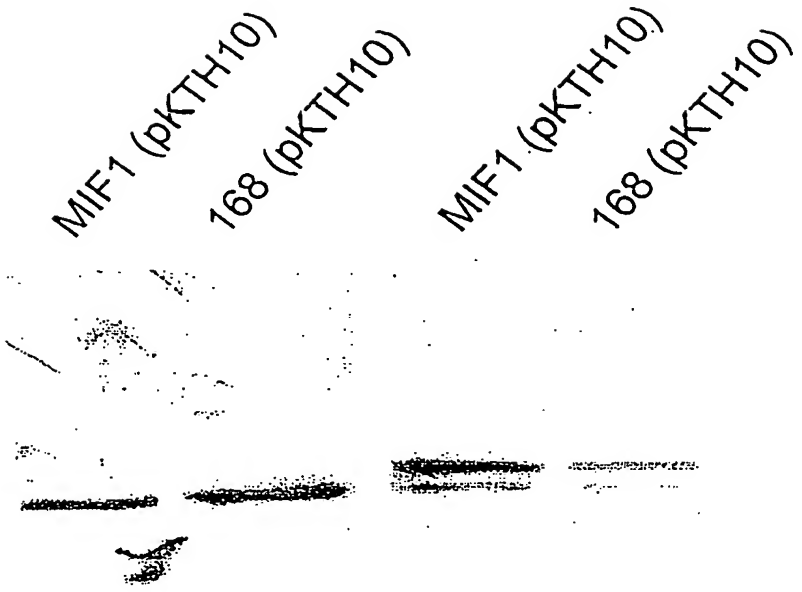
GC 385 Figure 1C

```
90                  1510                  1530
ACGACAGAGCAAGTTGAGAAGGATTTTGAATCTCTGGGTATGGACCCT
T  T  E  Q  V  E  K  D  F  E  S  L  G  M  D  P

          1550                  1570
GATACTGTAGTTCTGTCAGGCGAAAAGAGCAATATCGGTGTTGCCCGT
D  T  V  V  L  S  G  E  K  S  N  I  G  V  A  R

  1590                  1610                  1630
TTTGTCGGGGTGCCAGATAAAGAAACCATTGCAAAAGTAAAAACGTAT
F  V  G  V  P  D  K  E  T  I  A  K  V  K  T  Y

                1650                  1670
TTTAAAGACAAATACGGATCTGATCCAAATGTCAGCACAGTTTCACCG
F  K  D  K  Y  G  S  D  P  N  V  S  T  V  S  P

      1690                  1710                  17
ACAGTCGGTAAGGAGCTGGCGAGAAATGCGCTGTACGCAGTTGCTATA
T  V  G  K  E  L  A  R  N  A  L  Y  A  V  A  I

30                  1750                  1770
GCTTCTATTGGCATCATTATTTACGTTTCAATCCGATTCGAATACAAA
A  S  I  G  I  I  I  Y  V  S  I  R  F  E  Y  K

          1790                  1810
ATGGCGATTGCTGCCATCGCCTCATTGCTATATGACGCATTCTTTATC
M  A  I  A  A  I  A  S  L  L  Y  D  A  F  F  I

  1830                  1850                  1870
GTCACGTTCTTCAGTATTACAAGGCTTGAGGTAGATGTTACATTCATC
V  T  F  F  S  I  T  R  L  E  V  D  V  T  F  I


      1890                  1910
GCGGCCATCTTGACGATAATCGGGTATTCCATTAACGATACAATCGTT
A  A  I  L  T  I  I  G  Y  S  I  N  D  T  I  V

      1930                  1950                  19
ACATTTGACAGGGTCCGCGAGCATATGAAAAAGCGTAAGCCGAAAACC
T  F  D  R  V  R  E  H  M  K  K  R  K  P  K  T
```

GC385 Figure 1D

```
70                  1990                  2010
TTTGCCGATCTGAACCATATTGTAAACCTGAGCCTGCAGCAAACCTTT
F  A  D  L  N  H  I  V  N  L  S  L  Q  Q  T  F

          2030                  2050
ACACGTTCAATTAACACTGTATTAACCGTTGTGATTGTTGTTGTGACA
T  R  S  I  N  T  V  L  T  V  V  I  V  V  V  T

  2070                  2090                  2110
TTGCTGATCTTTGGAGCATCTTCTATTACTAACTTCTCAATTGCTTTA
L  L  I  F  G  A  S  S  I  T  N  F  S  I  A  L

              2130                  2150
TTGGTCGGGCTGTTAACAGGCGTTTATTCTTCTCTATACATTGCCGCA
L  V  G  L  L  T  G  V  Y  S  S  L  Y  I  A  A

      2170                  2190                  22
CAAATTTGGCTTGCATGGAAAGGAAGAGAACTGAAAAAAGATTCGGCG
Q  I  W  L  A  W  K  G  R  E  L  K  K  D  S  A

10
CAATAA
Q  *
```

GC 385 Figure 1E



GC 385

Figure 2

Expression of *secDF*

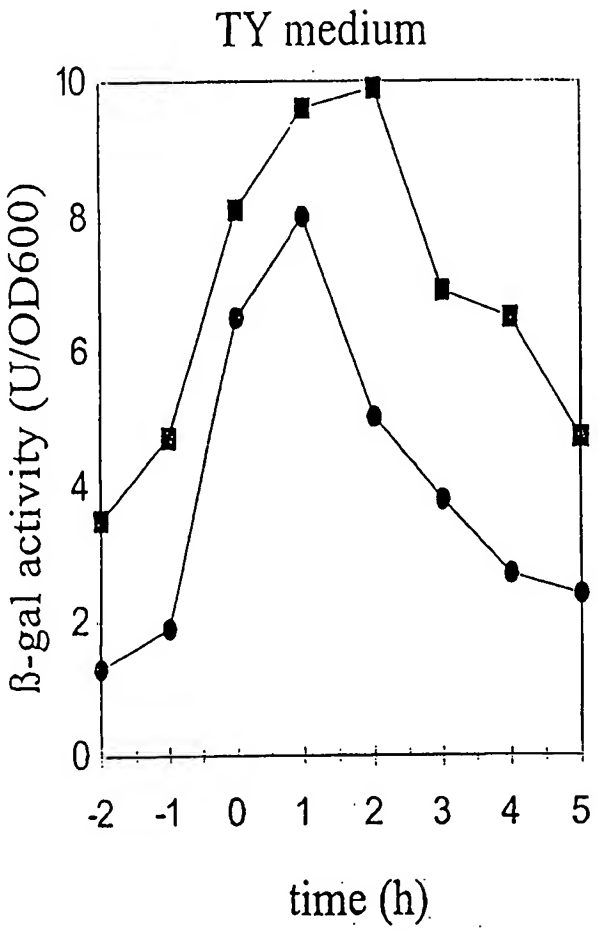


Figure 3A

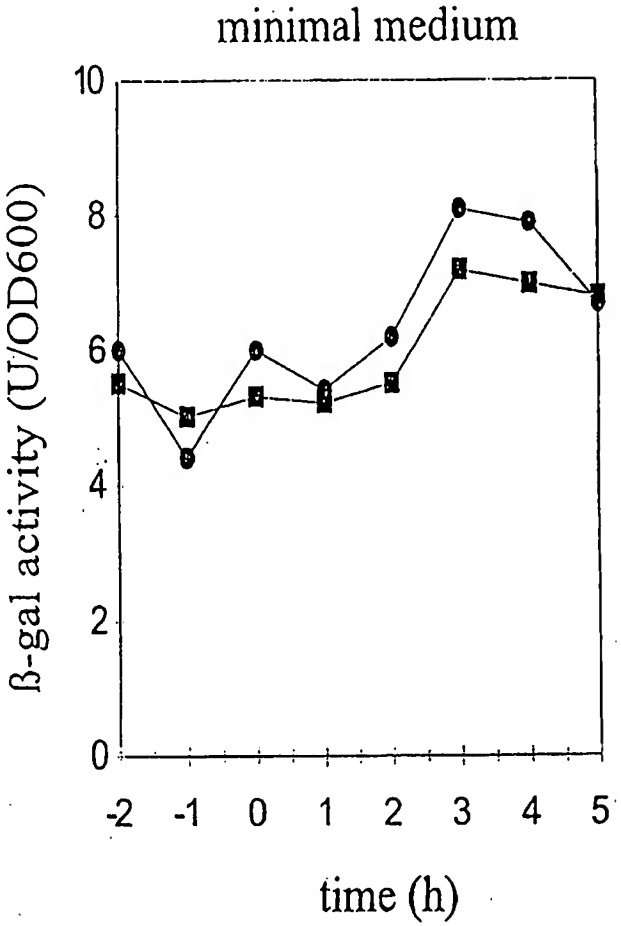
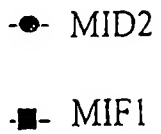


Figure 3B



GC385

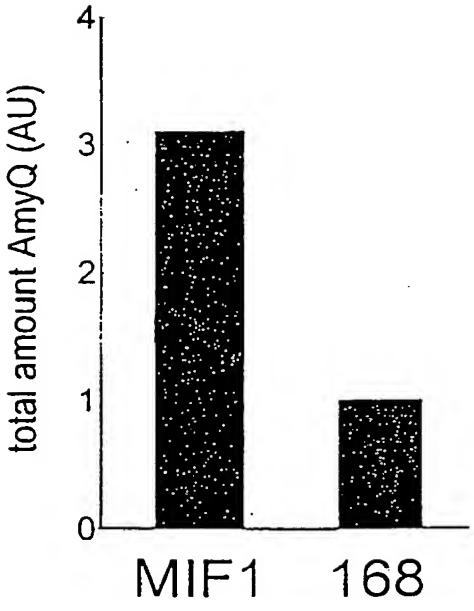


Figure 4A

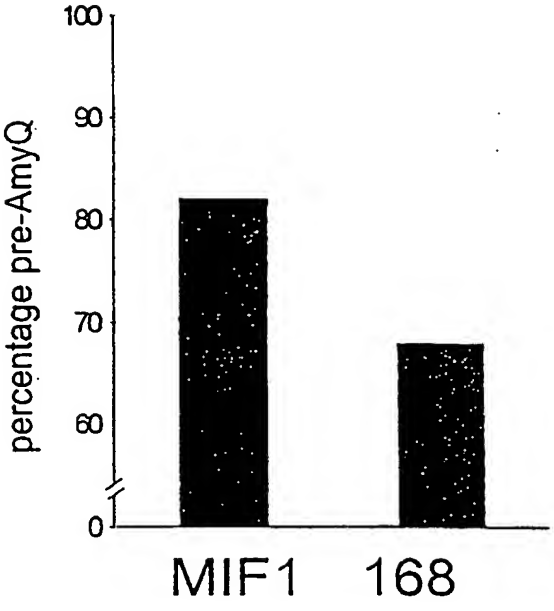


Figure 4B

GC 385

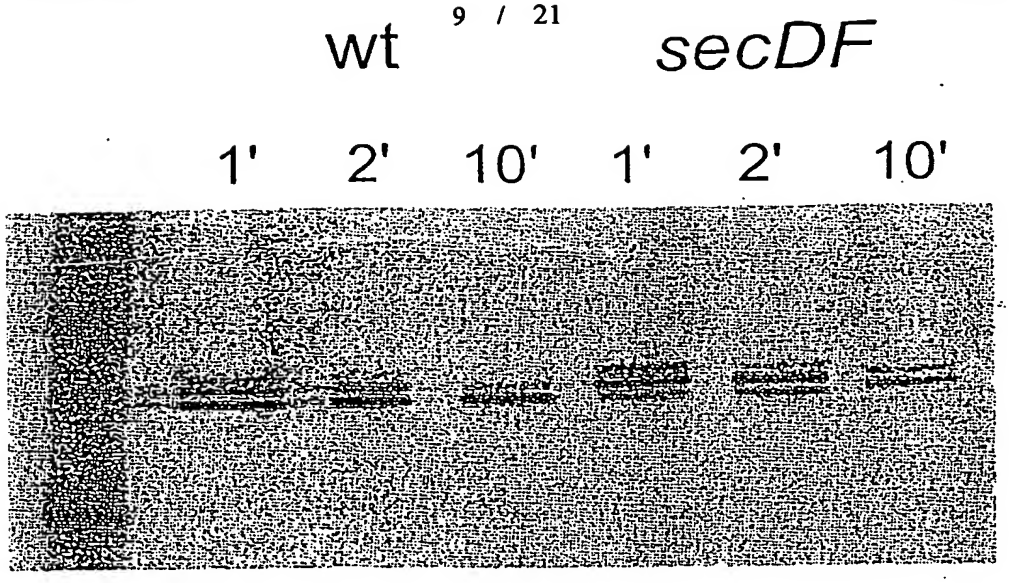


Figure 5A

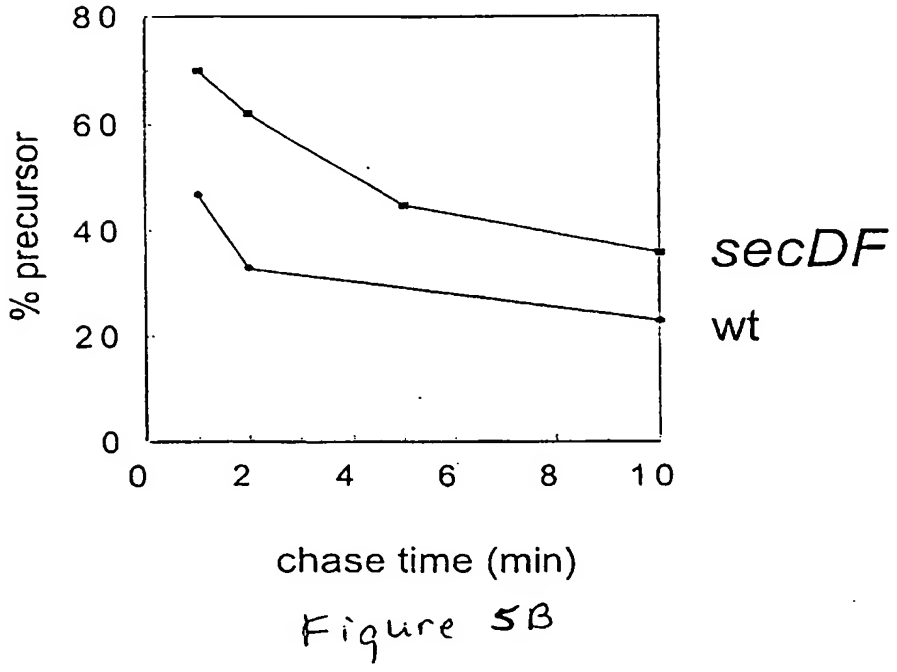
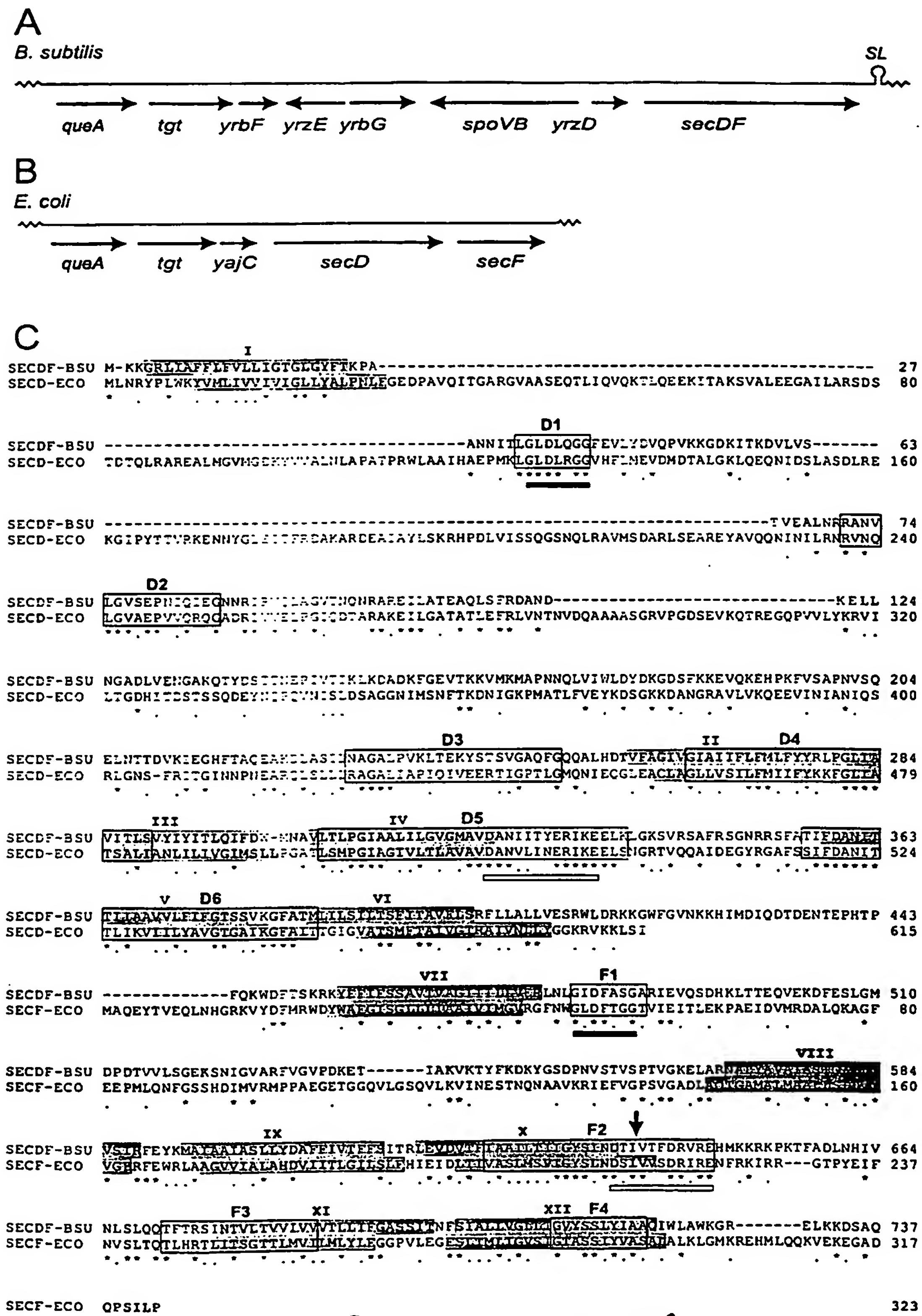


Figure 5B

GC 385



Figures 6A-6C

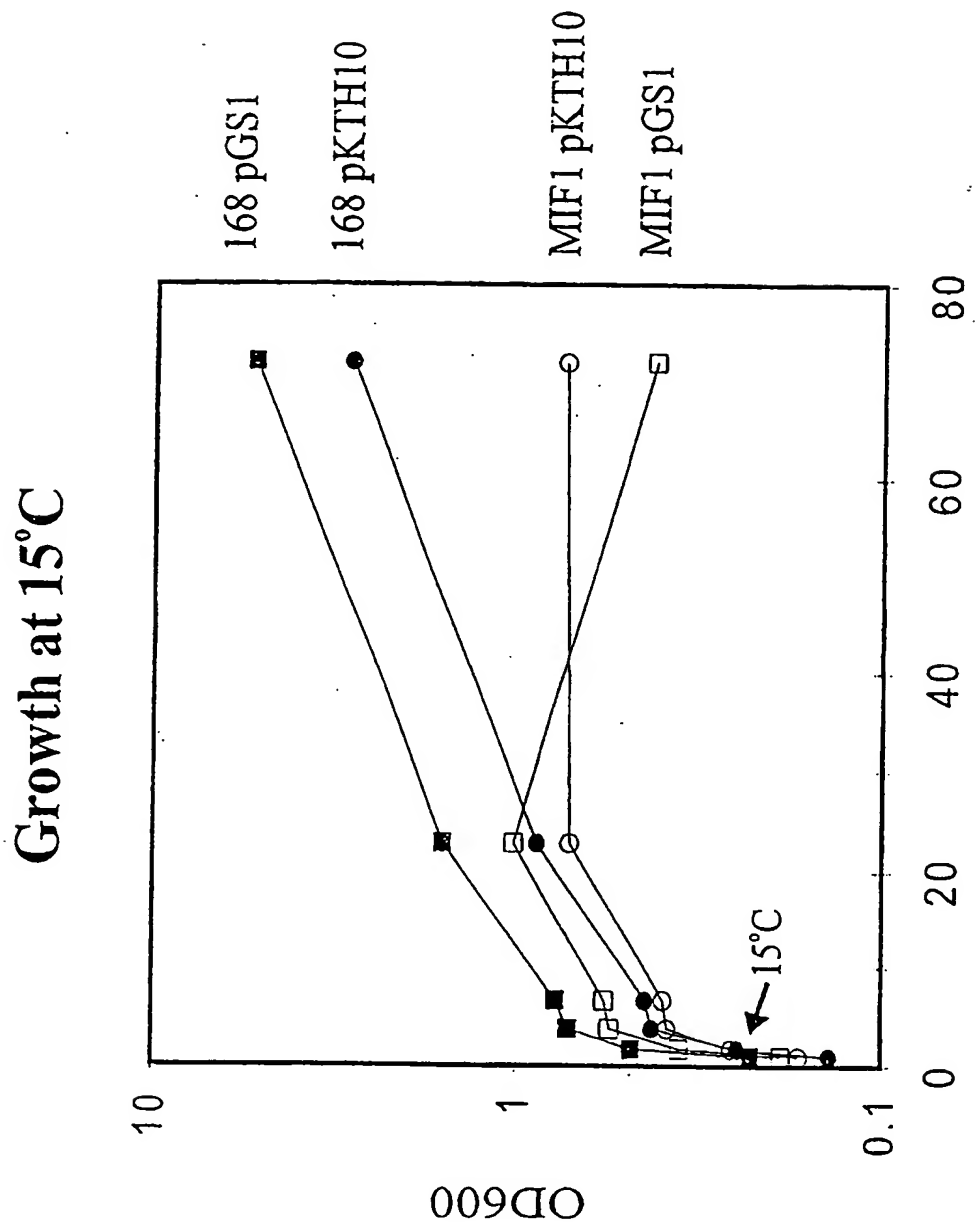
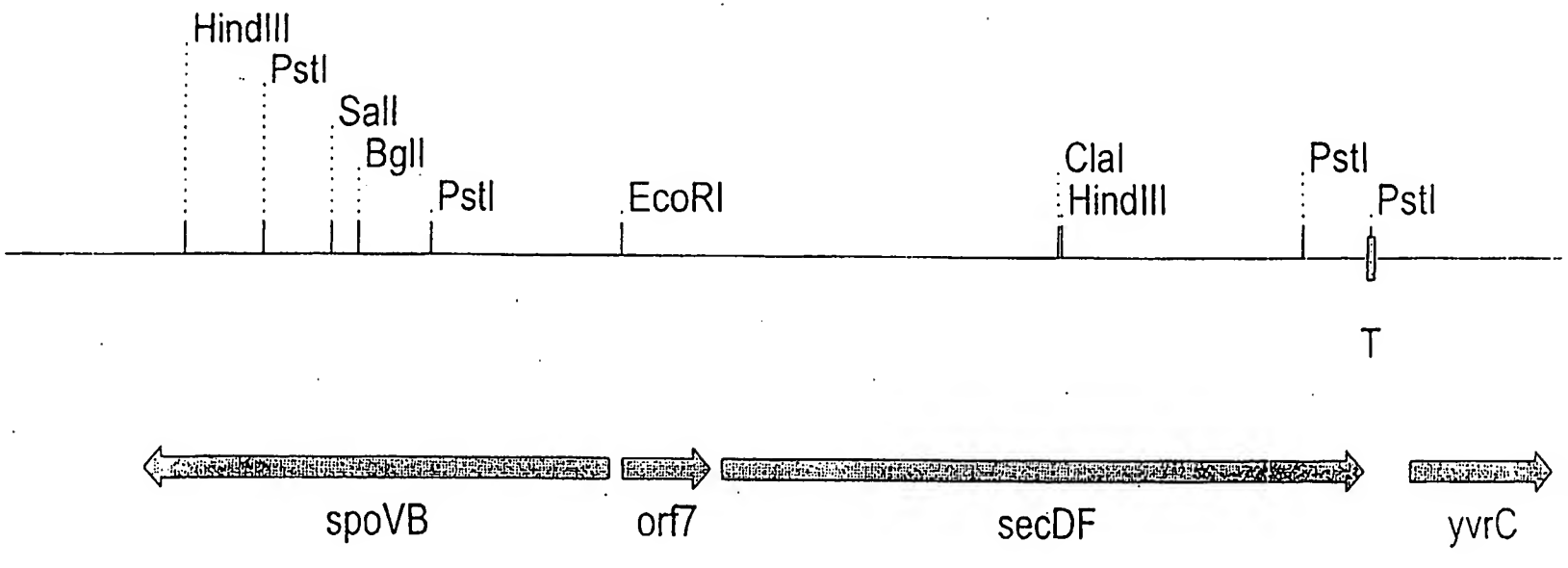


Figure 7

GC 385



(5459 bps)

Figure 8

GC 385

Cambell-integration pMUTIN2

5'- *secDF* fragment with RBS

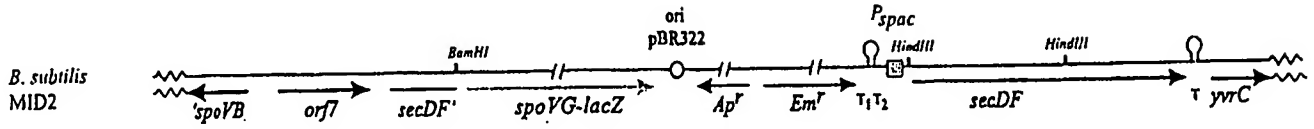


Figure 9A

internal *secDF* fragment

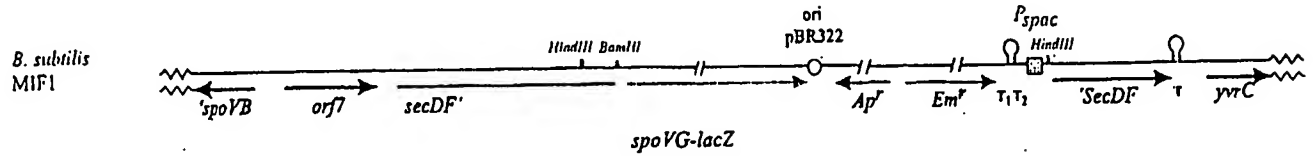
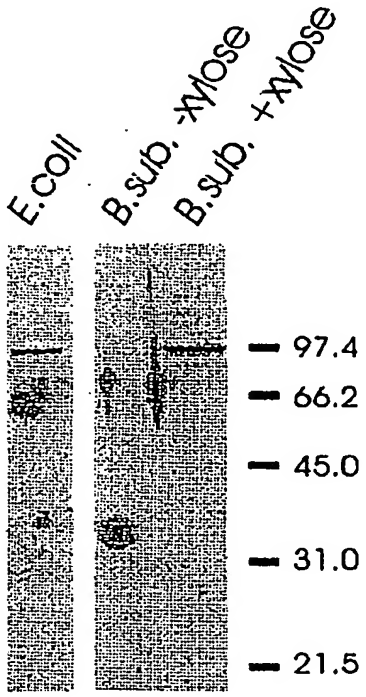


Figure 9B

GC 385

Expression of SecDF-myc



GC 385

Figure 1D

Secretion of Amylase (AmyQ)

1h; 37°C 16h; 15°C

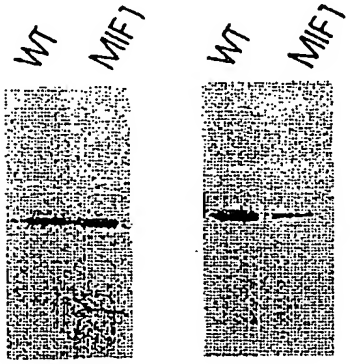


Figure 11

GC 385

Secretion of Neutral protease (NprE)

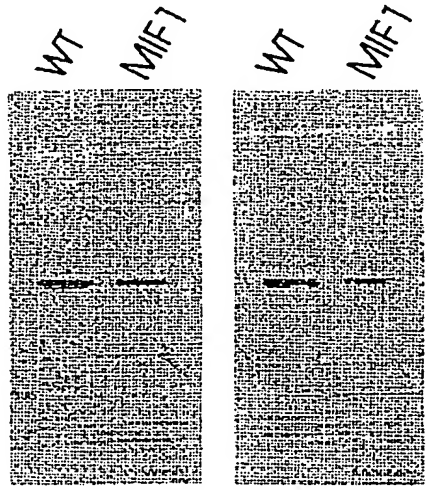
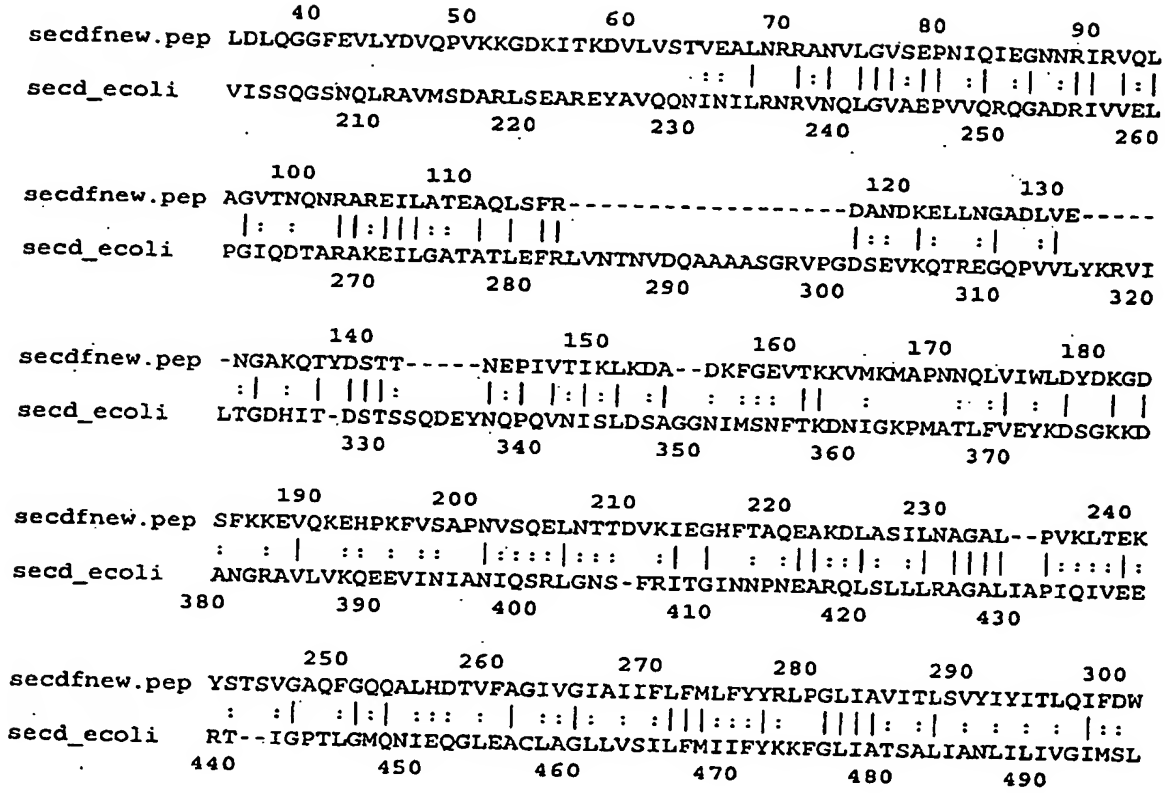


Figure 12

GC 385



GC 385 Figure 13

```
                250        260        270        280        290        300
secDFbsupep     LDRKKGWFGVNKKHIMDIQDTDENTEPHTPFQKWDFTSKRKYFFIFSSAVTVAGIIILLV
                                                |::||:     : | :|: : :|:|:|: |
SECF_ECOLI                    MAQEYTVEQLNHGRKVYDFMRWDY-----WAFGISGLLLIAAIVIMGV
                                        10        20             30        40

                310        320        330        340        350        360
secDFbsupep     FRLNLGIDFASGARIEVQSDHKLTTEQVEKDFESLGMDPDTVVLSGEKSNIGVARFVGVP
                  :| |:||::|: ||:  ::    : ::  ::: |::   :   | : :|    :| :|
SECF_ECOLI      RGFNWGLDFTGGTVIEITLEKPAEIDVMRDALQKAGFEEPMLQNFGSSHDI----MVRMP
                     50         60         70         80         90

                370                   380        390        400        410
secDFbsupep     DKE----------TIAKVKTYFKDKYGSDPNVSTVSPTVGKELARNALYAVAIASIGIII
                  |           : || :   :: ::   :  |:|:|| :||::: :|:  | ::|::
SECF_ECOLI      PAEGETGGQVLGSQVLKVINESTNQNAAVKRIEFVGPSVGADLAQTGAMALMAALLSILV
              100        110        120        130        140        150

                420        430        440        450        460        470
secDFbsupep     YVSIRFEYKMAIAAIASLLYDAFFIVTFFSITRLEVDVTFIAAILTIIGYSINDTIVTFD
                ||:|||:::| ::: :| :|::: : ::|: ::|:|:|::|:::::||||:||:||: |
SECF_ECOLI      YVGFRFEWRLAAGVVIALAHDVIITLGILSLFHIEIDLTIVASLMSVIGYSLNDSIVVSD
              160        170        180        190        200        210

                480        490        500        510        520        530
secDFbsupep     RVREHMKK-RKPKTFADLNHIVNLSLQQTFTRSINTVLTVVIVVVTLLIFGASSITNFSI
                |:||:::| |:   :    :| |:|| ||: |:: |  |:::|:: | :||:  : :||:
SECF_ECOLI      RIRENFRKIRRGTPY----EIFNVSLTQTLHRTLITSGTTLMVILMLYLFGGPVLEGFSL
              220        230            240        250        260        270

                 540        550        560        570
secDFbsupep     ALLVGLLTGVYSSLYIAAQIWLAWKGRELKKDSAQ
                ::|:|:  |: ||:|:|: : |
SECF_ECOLI      TMLIGVSIGTASSIYVASALALKLGMKREHMLQQKVEKEGADQPSILP
                  280        290        300        310        320
```

Figure 14

GC385

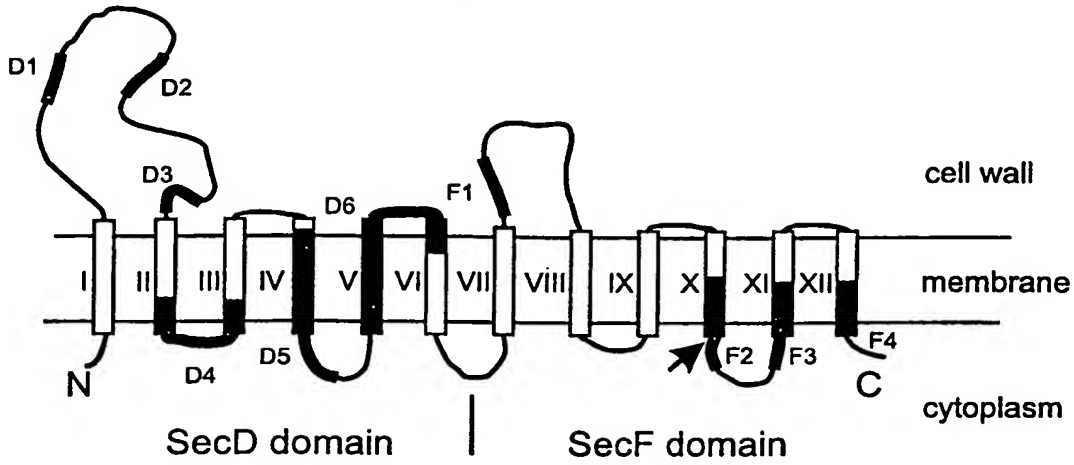
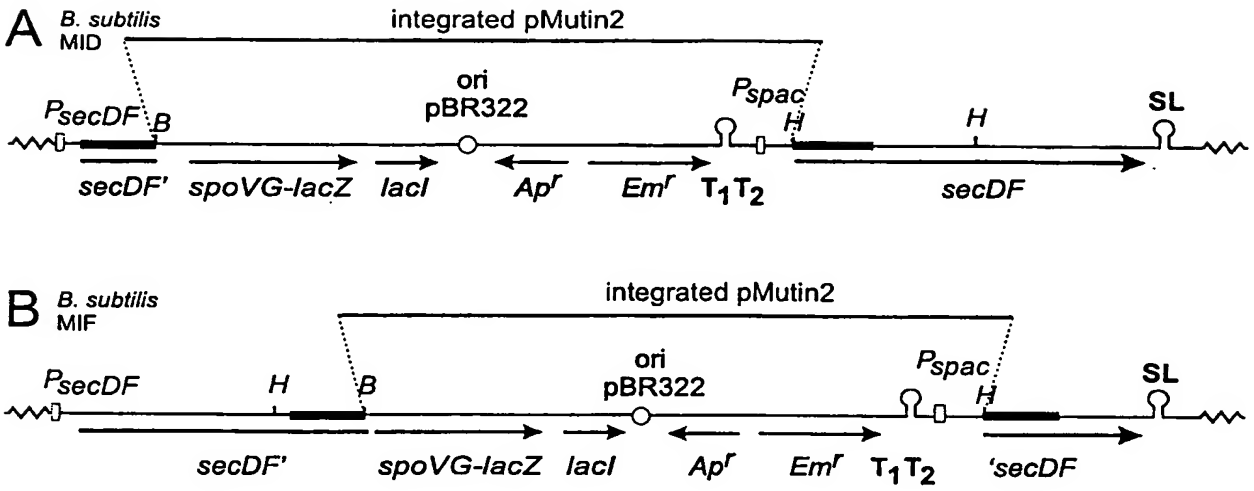

D1
D2
D3
D4
D5
D6
F1
F2
F3
F4
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
N
C
cell wall
membrane
cytoplasm
SecD domain
SecF domain

Figure 15



C

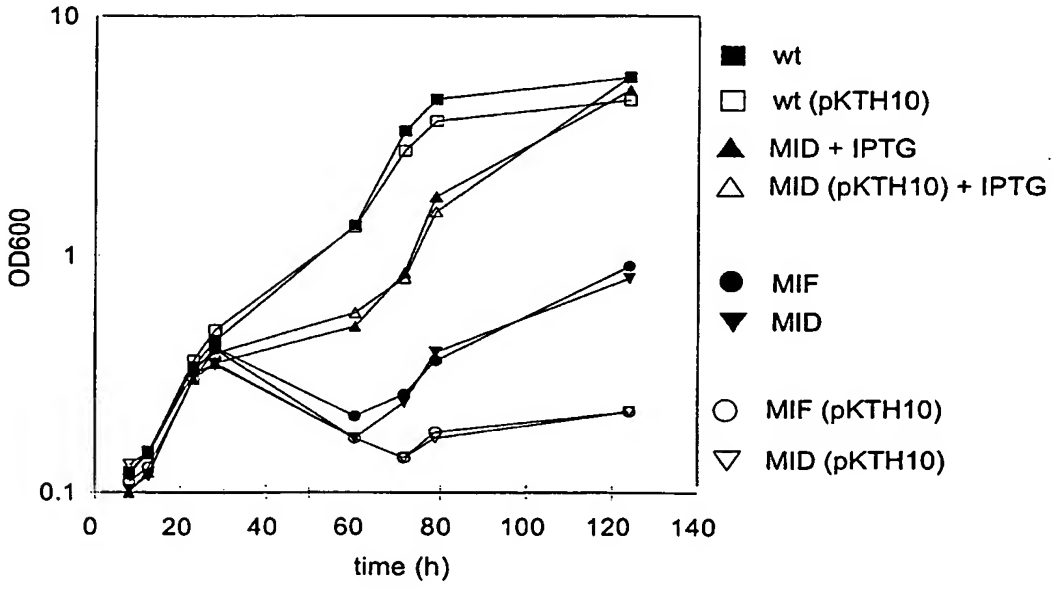


Figure 1A-1C

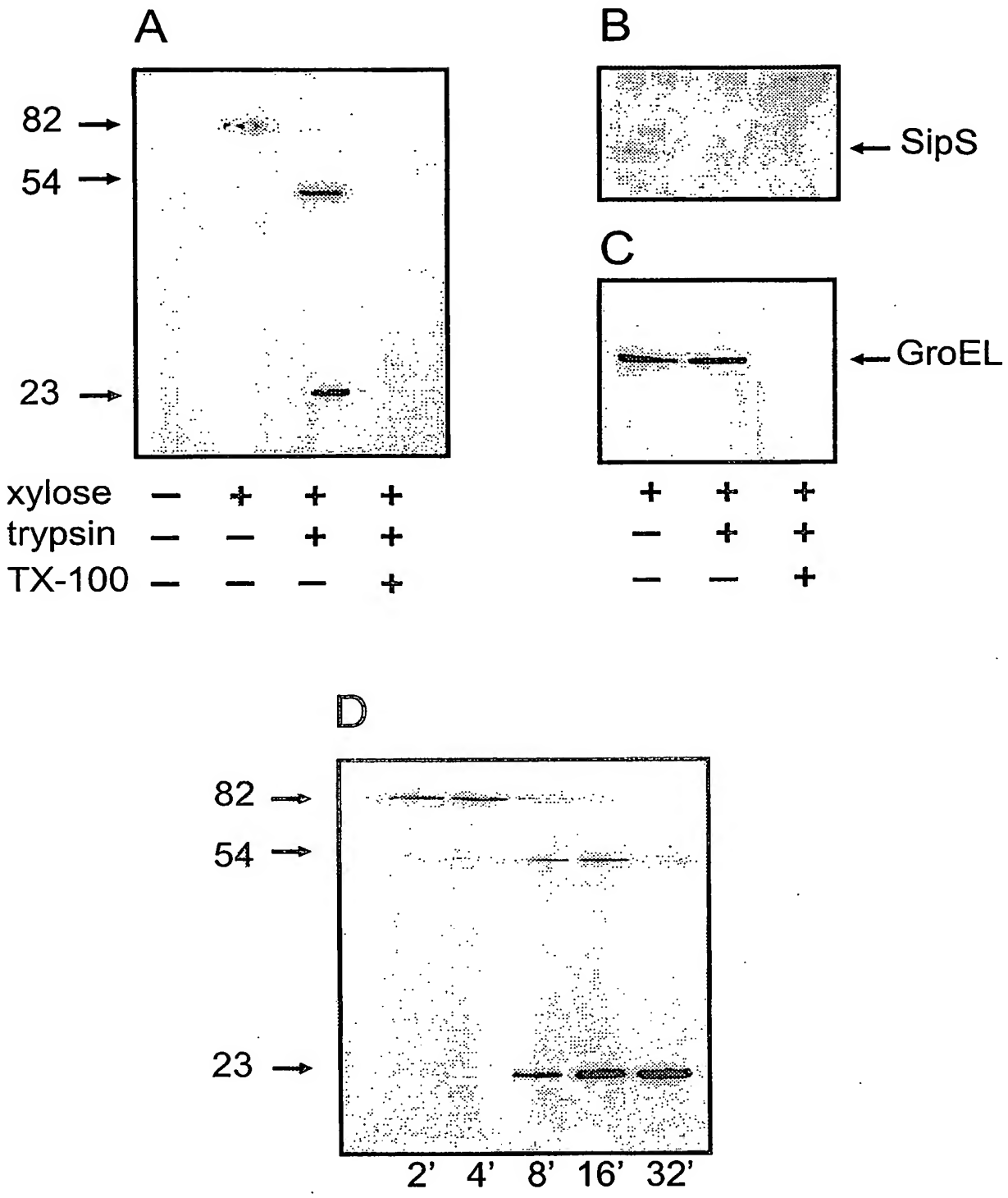


Figure 17A - 17D

INTERNATIONAL SEARCH REPORT

Int. dional Application No
PCT/US 98/14786

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/31 C07K14/32 C12N15/62 C12N15/75 //C12N1/21, C12R1/07

According to International Patent Classification (IPC) or to both national classification and IPC .

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GARDEL C ET AL: "The secD locus of E.coli codes for two membrane proteins required for protein export" EMBO JOURNAL, vol. 9, no. 10, 1990, pages 3209-3216, XP002084738 see abstract; figures 3,6 see page 3213, paragraph 3 - page 3215, paragraph 5 --- -/--	1-18

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 November 1998	04/12/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Oderwald, H

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int itional Application No
PCT/US 98/14786

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SUH J -W ET AL: "ISOLATION OF A SECY HOMOLOGUE FROM BACILLUS SUBTILIS: EVIDENCE FOR A COMMON PROTEIN EXPORT PATHWAY IN EUBACTERIA" MOLECULAR MICROBIOLOGY, vol. 4, no. 2, 1 January 1990, pages 305-314, XP000607144 see abstract; figure 3 see page 310, paragraph 3 see page 312, paragraph 5 ---	1-18
Y	WO 94 19471 A (FINNISH NAT PUBLIC HEALTH ;KONTINEN VESA (FI); SARVAS MATTI (FI)) 1 September 1994 see the claims see abstract; examples 2-5; table 3 see page 1, line 18 - page 2, line 11 see page 3, line 1 - page 7, line 21 see page 8, paragraph 2 see page 9, paragraph 3 - page 10, paragraph 1 ---	1-18
T	BOLHUIS A ET AL: "SecDF of Bacillus subtilis, a molecular siamese twin required for the efficient secretion of proteins" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 273, no. 33, 14 August 1998, pages 21217-21224, XP002084739 see the whole document -----	1-18

1

INTERNATIONAL SEARCH REPORT

.ernational application No.
PCT/US 98/ 14786

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims : 1-18 all partially

   An expression vector containing a gram-positiv SecD secretion factor encoding nucleic acid, a gram-positiv host bacterium transformed with said vector, a method for secretion of a protein comprising SecD.

2. Claims : 1-18 all partially

   as in invention 1 but comprising SecF.

3. Claims : 1-18 all partially

   as invention 1 but comprising SecDF.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int tional Application No
PCT/US 98/14786

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9419471 A	01-09-1994	AU 6142694 A	14-09-1994
		CA 2156425 A	01-09-1994
		CN 1120853 A	17-04-1996
		EP 0686195 A	13-12-1995
		FI 954010 A	25-10-1995
		JP 8507207 T	06-08-1996
		US 5780261 A	14-07-1998

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				